REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
30-09-2014	Final Report	2-Jun-2008 - 30-Jun-2014

4. TITLE AND SUBTITLE
Final Report: A Unified Approach to Abductive Inference

5a. CONTRACT NUMBER
W911NF-08-1-0242

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611103

6. AUTHORS
Pedro Domingos, Carlos Guestrin, Raymond Mooney, Thomas Dietterich, Henry Kautz, Joshua Tenenbaum, Daniel Lowd, Vibhav Gogate, Mathias Niepert

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Washington
Office of Sponsored Programs
4333 Brooklyn Avenue NE
Seattle, WA 98195 -9472

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
54231-NS-MUR.183

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The project's main focus was on tractable inference and learning of probabilistic representations, which are essential for large-scale abductive inference applications. We also developed novel inference techniques based on lifting, sampling, and more efficient processing of evidence. We continued to extend Alchemy 2.0, an open-source toolkit for Markov logic, and Alchemy Lite, an implementation of Tractable Markov Logic (TML). We developed parameter and structure learning algorithms for sum-product networks and, building on TML, we substantially

15. SUBJECT TERMS
Markov Logic, Abductive Inference, Tractable Probabilistic Knowledge Bases, Lifted Probabilistic Inference, Symmetry-based Learning and Inference, Non-convex Optimization, Combining Markov Logic and Support Vector Machines, Textual Inference,

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Pedro Domingos
					19b. TELEPHONE NUMBER
					206-543-4229

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Documentation Page					*Form Approved* *OMB No. 0704-0188*	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.						
1. REPORT DATE **2014**		2. REPORT TYPE **N/A**			3. DATES COVERED **-**	
4. TITLE AND SUBTITLE **Automatic Construction and Natural-Language Description of Nonparametric Regression Models**					5a. CONTRACT NUMBER	
					5b. GRANT NUMBER	
					5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)					5d. PROJECT NUMBER	
					5e. TASK NUMBER	
					5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Washington Office of Sponsored Programs 4333 Brooklyn Avenue NE Seattle, WA 98195 -9472**					8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					10. SPONSOR/MONITOR’S ACRONYM(S)	
					11. SPONSOR/MONITOR’S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT **The projects main focus was on tractable inference and learning of probabilistic representations, which are essential for large-scale abductive inference applications. We also developed novel inference techniques based on lifting, sampling, and more efficient processing of evidence. We continued to extend Alchemy 2.0, an open-source toolkit for Markov logic, and Alchemy Lite, an implementation of Tractable Markov Logic (TML). We developed parameter and structure learning algorithms for sum-product networks and, building on TML, we substantially improved two tractable probabilistic-logical formalisms: relational sum-product networks and tractable probabilistic knowledge bases. Based on sum-product networks, we worked towards formalisms for tractable probabilistic programming. We worked on symmetrybased inference and learning and developed novel model classes that exploit invariances of the data with respect to group operations. A novel model for Biomedical event extraction based on MLNs that leverages the power of support vector machines (SVMs) to handle highdimensional features was proposed and applied to the problem of event extraction. We developed structured prediction models by introducing novel forms of regularization. We continued to apply Markov logic networks to the problem of textual inference and conducted extensive experiments on benchmark datasets. We further improved GraphLab, our large-scale parallel machine learning framework. We investigated novel approaches to activity and plan recognition, and showed that Markov logic is capable of fusing visual and language evidence of the activities under consideration.**						
15. SUBJECT TERMS						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**63**		

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Report Title

Final Report: A Unified Approach to Abductive Inference

ABSTRACT

The project's main focus was on tractable inference and learning of probabilistic representations, which are essential for large-scale abductive inference applications. We also developed novel inference techniques based on lifting, sampling, and more efficient processing of evidence. We continued to extend Alchemy 2.0, an open-source toolkit for Markov logic, and Alchemy Lite, an implementation of Tractable Markov Logic (TML). We developed parameter and structure learning algorithms for sum-product networks and, building on TML, we substantially improved two tractable probabilistic-logical formalisms: relational sum-product networks and tractable probabilistic knowledge bases. Based on sum-product networks, we worked towards formalisms for tractable probabilistic programming. We worked on symmetry-based inference and learning and developed novel model classes that exploit invariances of the data with respect to group operations. A novel model for Biomedical event extraction based on MLNs that leverages the power of support vector machines (SVMs) to handle high-dimensional features was proposed and applied to the problem of event extraction. We developed structured prediction models by introducing novel forms of regularization. We continued to apply Markov logic networks to the problem of textual inference and conducted extensive experiments on benchmark datasets. We further improved GraphLab, our large-scale parallel machine learning framework. We investigated novel approaches to activity and plan recognition, and showed that Markov logic is capable of fusing visual and language evidence of the activities under consideration.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
05/05/2010	23.00	Patrick Shafto, Charles Kemp, Elizabeth Baraff Bonawitz John D. Coley, Joshua B. Tenenbaum. Inductive reasoning about causally transmitted properties, Cognition, (01 2008): . doi:
05/05/2010	24.00	Thomas L. Griffiths, Joshua B. Tenenbaum. Theory-Based Causal Induction, Psychological Review, (01 2009): . doi:
06/24/2009	11.00	Liangliang Cao, Jiebo Luo, Henry Kautz, Thomas S. Huang. Image Annotation Within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models, IEEE Transactions on Multimedia, (2009): . doi:
08/28/2012	04.00	Edward Vul, Joshua B. Tenenbaum, Samuel J. Gershman. Multistability and Perceptual Inference, Neural Computation, (01 2012): 0. doi: 10.1162/NECO_a_00226
08/28/2013	29.00	Noah Goodman, Tomer Ullman, Joshua Tenenbaum. Theory learning as stochastic search in the language of thought, Cognitive Devolopment, (10 2012): 455. doi: 10.1016/j.cogdev.2012.07.005
08/29/2013	37.00	Jonathan Huang, Ashish Kapoor, Carlos Guestrin. Riffled Independence for Efficient Inference with Partial Rankings, Journal of Artificial Intelligence Research, (07 2012): 491. doi:
08/29/2013	38.00	Dafna Shahaf, Carlos Guestrin. Connecting Two (or Less) Dots: Discovering Structure in News Articles, ACM Transactions on Knowledge Discovery from Data, (02 2012): 24. doi: 10.1145/2086737.2086744
08/31/2011	71.00	J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman. How to Grow a Mind: Statistics, Structure, and Abstraction, Science, (03 2011): 0. doi: 10.1126/science.1192788
08/31/2011	72.00	E. Vul, V. Girotto, M. Gonzalez, J. B. Tenenbaum, L. L. Bonatti, E. Teglas. Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference, Science, (05 2011): 0. doi: 10.1126/science.1196404
09/03/2014	57.00	Daniel Lowd, Jesse Davis. Improving Markov Network Structure Learning Using Decision Trees, Journal of Machine Learning Research, (02 2014): 501. doi:
09/29/2014	77.00	Peter Battaglia, Jessica Hamrick, Joshua Tenenbaum. Simulation as an engine of physicalscene understanding, Proceedings of the National Academy of Sciences, (11 2013): 18327. doi:

TOTAL: 11

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received		Paper
08/24/2012	88.00	Adam Sadilek, Henry Kautz. Location-Based Reasoning about Complex Multi-Agent Behavior, Journal of Artificial Intelligence Research, (01 2012): 87. doi:
TOTAL:		**1**

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper
08/23/2013 13.00	Deepak Venugopal, Vibhav Gogate. Lifting WALKSAT-based Local Search Algorithms for MAP Inference, AAAI-13 Workshop on Statistical Relational Artificial Intelligence. 14-JUL-13, . : ,
08/25/2014 48.00	Yinon Bentor, Amelia Harrison, Shruti Bhosale, Raymond Mooney. Slot Filling System: Bayesian Logic Programs for Textual Inference, Sixth Text Analysis Conference (TAC 2013). 18-NOV-13, . : ,
08/27/2012 96.00	Daniel Lowd, Ali Torkamani. Towards Adversarial Collective Classification, ARO Workshop on Moving Target Defense. 01-OCT-11, . : ,
08/27/2012 03.00	Pedro Domingos, Chloe Kiddon. Knowledge Extraction and Joint Inference Using Tractable Markov Logic, Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction at NAACL-12. 07-JUN-12, . : ,
08/27/2013 21.00	Tivadar Papai, Henry Kautz. Modal Markov Logic for Multiple Agents, 3rd International Workshop on Statistical Relational AI. 14-JUL-13, . : ,
08/31/2011 87.00	Yi Chu, Young Chol Song, Henry Kautz, Richard Levinson. When Did You Start Doing That Thing That You Do? Interactive Activity Recognition and Prompting, AAAI 2011 Workshop on Artificial Intelligence and Smarter Living: The Conquest of Complexity. 08-AUG-11, . : ,
09/03/2014 65.00	Tivadar Papai, Henry Kautz. Modal Markov Logic for Multiple Agents, 3rd International Workshop on Statistical Relational AI. 15-JUL-13, . : ,
TOTAL:	**7**

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
08/24/2012	89.00	Henry Kautz, Adam Sadilek, Jeffrey P. Bigham. Finding your friends and following them to where you are, the fifth ACM international conference. 07-FEB-12, Seattle, Washington, USA. : ,
08/24/2012	94.00	Daniel Lowd. Closed-Form Learning of Markov Networks from Dependency Networks, Conference on Uncertainty in Artificial Intelligence. 15-AUG-12, . : ,
08/24/2012	92.00	Adam Sadilek, Henry Kautz, Vincent Silenzio. Modeling Spread of Disease from Social Interactions, International Conference on Weblogs and Social Media. 04-JUN-12, . : ,
08/24/2012	91.00	Adam Sadilek, Henry Kautz, Vincent Silenzio. Predicting Disease Transmission from Geo-Tagged Micro-Blog Data, AAAI Conference on Artificial Intelligence. 22-JUL-12, . : ,
08/24/2012	90.00	Tivadar Papai, Shalini Ghosh, Henry Kautz. Combining Subjective Probabilities and Data in Training Markov Logic Networks, European Conference on Machine Learning. 24-SEP-12, . : ,
08/26/2013	14.00	Deepak Venugopal, Vibhav Gogate. Dynamic Blocking and Collapsing for Gibbs Sampling, The 29th Conference on Uncertainty in Artificial Intelligence. 11-JUL-13, . : ,
08/26/2013	18.00	Brian King, Alan Fern, Jesse Hostetler. On Adversarial Policy Switching with Experiments in Real-Time Strategy Games, 23rd International Conference on Automated Planning and Scheduling. 10-JUN-13, . : ,
08/26/2013	17.00	Deepak Venugopal, Vibhav Gogate. GiSS: Combining SampleSearch and Importance Sampling for Inference in Mixed Probabilistic and Deterministic Graphical Models, The 27th AAAI Conference on Artificial Intelligence. 14-JUL-13, . : ,
08/26/2013	16.00	David Smith, Vibhav Gogate. The Inclusion-Exclusion Rule and its Application to the Junction Tree Algorithm, 23rd International Joint Conference on Artificial Intelligence. 03-AUG-13, . : ,
08/26/2013	15.00	Vibhav Gogate, Pedro Domingos. Structured Message Passing, The 29th Conference on Uncertainty in Artificial Intelligence. 11-JUL-13, . : ,
08/26/2014	49.00	Somdeb Sarkhel, Deepak Venugopal, Parag Singla, Vibhav Gogate. Lifted MAP Inference for Markov Logic Networks, 17th International Conference on Artificial Intelligence and Statistics (AISTATS), 2014.. 14-APR-14, . : ,
08/26/2014	52.00	Tahrima Rahman,, Prasanna Kothalkar, Vibhav Gogate. Cutset Networks: A Simple, Tractable, and Scalable Approach for Improving the Accuracy of Chow-Liu Trees, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2014.. 15-SEP-14, . : ,
08/26/2014	51.00	Deepak Venugopal, Vibhav Gogate. Evidence-Based Clustering for Scalable Inference in Markov Logic, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2014.. 15-SEP-14, . : ,
08/26/2014	50.00	David Smith, Vibhav Gogate. Loopy Belief Propagation in the Presence of Determinism, 17th International Conference on Artificial Intelligence and Statistics (AISTATS), 2014.. 22-APR-14, . : ,

08/27/2012 95.00 Ali Torkamani, Daniel Lowd. Convex Adversarial Collective Classification,
International Workshop on Statistical Relational AI . 15-AUG-12, . : ,

08/27/2012 02.00 Pedro Domingos, W. AustinWebb. Tractable Markov Logic,
Statistical Relational Learning workshop at ICML-12. 30-JUN-12, . : ,

08/27/2012 01.00 Pedro Domingos, Aniruddh Nath. Learning Multiple Hierarchical Relational Clusterings,
Statistical Relational Learning workshop at ICML-12. 30-JUN-12, . : ,

08/27/2012 00.00 Pedro Domingos, W. Austin Webb. A Tractable First-Order Probabilistic Logic,
AAAI Conference on Artificial Intelligence. 22-JUL-12, . : ,

08/27/2012 99.00 Sindhu Raghavan, Raymond J. Mooney, Hyeonseo Ku. Learning to "Read Between the Lines" using Bayesian Logic Programs,
Annual Meeting of the Association for Computational Linguistics. 08-JUL-12, . : ,

08/27/2012 98.00 Deepak Venugopal, Vibhav Gogate. On Lifting the Gibbs Sampling Algorithm,
Workshop on Statistical Relational AI. 15-AUG-12, . : ,

08/27/2012 97.00 Vibhav Gogate, Abhay Jha, Deepak Venugopal. Advances in Lifted Importance Sampling,
AAAI Conference on Artificial Intelligence. 22-JUL-12, . : ,

08/27/2013 19.00 Islam Beltagyx, Cuong Chaux, Gemma Boleday, Dan Garrettex, Katrin Erky, Raymond Mooney. Montague Meets Markov: Deep Semantics with Probabilistic Logical Form,
2nd Joint Conference on Lexical and Computational Semantics. 13-JUN-13, . : ,

08/27/2013 27.00 Daniel Lowd, Amirmohammad Rooshenas. Learning Markov Networks With Arithmetic Circuits,
30th International Conference on Machine Learning. 20-JUN-13, . : ,

08/27/2013 26.00 MohamadAli Torkamani, Daniel Lowd. Convex Adversarial Collective Classification,
30th International Conference on Machine Learning. 18-JUN-13, . : ,

08/27/2013 25.00 Daniel Lowd, Amirmohammad Rooshenas. Learning Markov Networks With Arithmetic Circuits,
16th International Conference on Artificial Intelligence and Statistics. 30-APR-13, . : ,

08/27/2013 24.00 Tivadar Papai, Henry Kautz, Daniel Stefankovic. Slice Normalized Dynamic Markov Logic Networks,
Neural Information Processing Systems. 04-DEC-12, . : ,

08/27/2013 23.00 Tivadar Papai, Henry Kautz, Daniel Stefankovic. Reasoning Under the Principle of Maximum Entropy for Modal Logics K45, KD45, and S5,
14th Conference on Theoretical Aspects of Rationality and Knowledge. 08-JAN-13, . : ,

08/27/2013 22.00 Adam Sadilek, Henry Kautz. Modeling the Impact of Lifestyle on Health at Scale,
6th ACM International Conference on Web Search and Data Mining. 06-FEB-13, . : ,

08/27/2013 20.00 Sindhu Raghavan, Raymond Mooney. Online Inference-Rule Learning from Natural-Language Extractions,
AAAI-13 Workshop on Statistical Relational AI. 14-JUL-13, . : ,

08/27/2014 53.00 Deepak Venugopal , Vibhav Gogate, Chen Chen, Vincent Ng. Relieving the Computational Bottleneck: Joint Inference for Event Extraction with High-Dimensional Features,
2014 Conference on Empirical Methods on Natural Language Processing, (EMNLP), 2014. 08-OCT-14, . : ,

08/28/2012 05.00 Joseph K. Bradley, Carlos Guestrin. Sample Complexity of Composite Likelihood,
Conference on Artificial Intelligence and Statistics. 23-APR-12, . : ,

08/28/2012 10.00 Ethan Dereszynski, Jesse Hostetler, Tom Dietterich, Alan Fern. Inferring Strategies from Limited Reconnaissance in Real-time Strategy Games,
Conference on Uncertainty in Artificial Intelligence. 15-AUG-12, . : ,

08/28/2012 09.00 Jonathan Huang, Ashish Kapoor, Carlos Guestrin. Efficient Probabilistic Inference with Partial Ranking Queries,
Conference on Uncertainty in Artificial Intelligence. 14-JUL-11, . : ,

08/28/2012 08.00 Carlos Guestrin, Eric Horvitz, Dafna Shahaf. Metro maps of science,
the 18th ACM SIGKDD international conference. 11-AUG-12, Beijing, China. : ,

08/28/2012 07.00 Joseph Gonzalez, Aapo Kyrola, Yucheng Low, Danny Bickson, Carlos Guestrin, Joseph M. Hellerstein. Distributed GraphLab: A Framework for Machine Learningand Data Mining in the Cloud,
Conference on Very Large Databases. 27-AUG-12, . : ,

08/28/2012 06.00 Dafna Shahaf, Carlos Guestrin, Eric Horvitz. Trains of thought: generating information maps,
World Wide Web Conference. 15-APR-12, Lyon, France. : ,

08/28/2013 30.00 Julian Jara-Ettinger, Chris Baker, Joshua Tenenbaum. Learning What is Where from Social Observations,
34th Annual Conference of the Cognitive Science Society. 01-AUG-12, . : ,

08/28/2013 36.00 William Webb, Pedro Domingos. Tractable Probabilistic Knowledge Bases with Existence Uncertainty,
3rd International Workshop on Statistical Relational AI. 14-JUL-13, . : ,

08/28/2013 35.00 Robert Gens, Pedro Domingos. Discriminative Learning of Sum-Product Networks,
Conference on Neural Information Processing Systems. 03-DEC-12, . : ,

08/28/2013 34.00 Robert Gens, Pedro Domingos. Learning the Structure of Sum-Product Networks,
30th International Conference on Machine Learning. 16-JUN-13, . : ,

08/28/2013 32.00 Eyal Dechter, Jon Malmaud, Ryan Adams, Joshua Tenenbaum. Bootstrap Learning via Modular Concept Discovery,
23rd International Joint Conference on Artificial Intelligence. 03-AUG-13, . : ,

08/28/2013 31.00 Roger Grosse, Ruslan Salakhutdinov, William Freeman, Joshua Tenenbaum. Exploiting compositionality to explore a large space of model structures,
28th Conference on Uncertainty in Artificial Intelligence. 15-AUG-12, . : ,

08/28/2014 54.00 Vibhav Gogate, Pedro Domingos. Structured Message Passing,
Twenty-Ninth Conference on Uncertainty in Artificial Intelligence. 12-JUL-13, . : ,

08/28/2014 56.00 Mathias Niepert, Pedro Domingos. Exchangeable Variable Models,
Thirty-First International Conference on Machine Learning. 21-JUN-14, . : ,

08/28/2014 55.00 Parag Singla, Aniruddh Nath, Pedro Domingos. Approximate Lifting Techniques for Belief Propagation,
Twenty-Eighth AAAI Conference on Artificial Intelligence. 27-JUL-14, . : ,

08/29/2013 39.00 Aapo Kyrola, Guy Blelloch, Carlos Guestrin. GraphChi: Large-Scale Graph Computation on Just a PC,
10th USENIX Symposium on Operating Systems Design and Implementation. 08-OCT-12, . : ,

08/29/2013 40.00 Joseph Gonzalez, Yucheng Low, Haijie Gu, Danny Bickson, Carlos Guestrin. PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs,
10th USENIX Symposium on Operating Systems Design and Implementation. 08-OCT-12, . : ,

08/31/2011 73.00 Chris L. Baker, Rebecca R. Saxe, Joshua B. Tenenbaum. Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution,
Thirty-Third Annual Conference of the Cognitive Science Society. 20-JUL-11, . : ,

08/31/2011 86.00 Raymond J. Mooney, Parag Singla. Abductive Markov Logic for Plan Recognition,
Conference on Artificial Intelligence. 07-AUG-11, . : ,

08/31/2011 85.00 Sindhu Raghavan, Raymond J. Mooney. Abductive Plan Recognition by Extending Bayesian Logic Programs,
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. 05-SEP-11, . : ,

08/31/2011 84.00 Hoifung Poon, Pedro Domingos. Sum-Product Networks: A new Deep Architecture,
Conference on Uncertainty in Artificial Intelligence. 14-JUL-11, . : ,

08/31/2011 83.00 Parag Singla, Henry Kautz, Tivadar Papai. Constraint Propagation for Efficient Inference inMarkov Logic,
International Conference on Principles and Practice of Constraint Programming. 12-SEP-11, . : ,

08/31/2011 82.00 Chloe Kiddon, Pedro Domingos. Coarse-to-fine Inference for First-order Probabilisticmodels,
Conference on Artificial Intelligence. 07-AUG-11, . : ,

08/31/2011 81.00 Tuyen N. Huynh, Raymond J. Mooney. Online Max-Margin Weight Learning for Markov Logic Networks,
SIAM International Conference on Data Mining. 28-APR-11, . : ,

08/31/2011 80.00 Tuyen N. Huynh, Raymond J. Mooney. Online Structure Learning forMarkov Logic Networks,
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. 05-SEP-11, . : ,

08/31/2011 79.00 Arthur Gretton, Carlos Guestrin, Yucheng Low, Joseph E. Gonzalez. Parallel Gibbs Sampling: From Colored Fields to Thin Junction Trees,
International Conference on Artificial Intelligence and Statistics. 11-APR-11, . : ,

08/31/2011 78.00 Vibhav Gogate, Pedro Domingos. Probabilistic Theorem Proving,
Conference on Uncertainty in Artificial Intelligence. 14-JUL-11, . : ,

08/31/2011 77.00 Vibhav Gogate, Pedro Domingos. Approximation by Quantization,
Conference on Uncertainty in Artificial Intelligence. 14-JUL-11, . : ,

08/31/2011 76.00 Ethan Dereszynski, Jesse Hostetler, Alan Fern, Tom Dietterich, Mark Udarbe, Thao-Trang Hoang. Learning Probabilistic Behavior Models in Real-time Strategy Games,
Artificial Intelligence in Digital Entertainment Conference. 14-OCT-11, . : ,

08/31/2011 75.00 Joseph K. Bradley, Aapo Kyrola, Danny Bickson, Carlos Guestrin. Parallel Coordinate Descent for L1-Regularized Loss Minimization,
International Conference on Machine Learning. 28-JUN-11, . : ,

08/31/2011 74.00 James Blythe, Jerry R. Hobbs, Raymond J. Mooney, Pedro Domingos, Rohit J. Kate. Implementing Weighted Abduction in Markov Logic,
International Conference on Computational Semantics. 12-JAN-11, . : ,

09/03/2014 58.00 Mohamad Ali Torkamani, Daniel Lowd. On Robustness and Regularization of Structural Support Vector Machines,
31st International Conference on Machine Learning (ICML). 21-JUN-14, . : ,

09/03/2014 64.00 Sean Brennan, Adam Sadilek, Henry Kautz. Towards Understanding Global Spread of Disease from Everyday Interpersonal Interactions,
23rd International Joint Conference on Artificial Intelligenced (IJCAI). 03-AUG-13, . : ,

09/03/2014 63.00 Adam Sadilek└, Sean Brennan, Henry Kautz, Vincent Silenzio. nEmesis: Which Restaurants Should You Avoid Today?,
First AAAI Conference on Human Computation and Crowdsourcing (HCOMP). 07-NOV-13, . : ,

09/03/2014 62.00 Young Chol Song, Henry Kautz, James Allen, Mary Swift, Yuncheng Li, Jiebo Luo. A Markov Logic Framework for Recognizing Complex Events from Multimodal Data,
15th ACM International Conference on Multimodal Interaction. 09-DEC-13, . : ,

09/03/2014 59.00 Amirmohammad Rooshenas, Daniel Lowd. Learning Sum-Product Networks with Direct and Indirect Variable Interactions,
31st International Conference on Machine Learning (ICML). 21-JUN-14, . : ,

09/11/2014 66.00 Aniruddh Nath, Pedro Domingos. Learning Tractable Statistical Relational Models,
Fourth International Workshop on Statistical Relational Artificial Intelligence. 22-JUL-14, . : ,

09/11/2014 70.00 Robert Peharz, Robert Gens, Pedro Domingos. Learning Selective Sum-Product Networks,
ICML-2014 Workshop on Learning Tractable Probabilistic Models. 21-JUN-14, . : ,

09/11/2014 69.00 Abram Friesen, Pedro Domingos. Exploiting Structure for Tractable Nonconvex Optimization,
ICML-2014 Workshop on Learning Tractable Probabilistic Models. 21-JUN-14, . : ,

09/11/2014 68.00 Aniruddh Nath, Pedro Domingos. Automated Debugging with Tractable Probabilistic Programming,
Fourth International Workshop on Statistical Relational Artificial Intelligence. 22-JUL-14, . : ,

09/11/2014 67.00 Aniruddh Nath, Pedro Domingos. Learning Tractable Statistical Relational Models,
ICML-2014 Workshop on Learning Tractable Probabilistic Models. 21-JUN-14, . : ,

09/22/2014 72.00 Mathias Niepert, Pedro Domingos. Tractable Probabilistic Knowledge Bases:Wikipedia and Beyond,
Fourth International Workshop on Statistical Relational Artificial Intelligence. 22-JUL-14, . : ,

09/22/2014 74.00 Mathias Niepert, Pedro Domingos. Exchangeable Variable Models,
ICML-2014 Workshop on Learning Tractable Probabilistic Models. 21-JUN-14, . : ,

09/22/2014 73.00 Mathias Niepert, Guy Van den Broeck. Tractability through Exchangeability: A New Perspective on Efficient Probabilistic Inference,
28th Conference on ArtificialIntelligence (AAAI). 27-JUL-14, . : ,

09/24/2014 76.00 Jesse Hostetler, Alan Fern, Tom Dietterich. State Aggregation in Monte Carlo Tree Search,
AAAI Conference on Artificial Intelligence. 29-JUL-14, . : ,

09/29/2014 81.00 Tobias Gerstenberg, Noah Goodman, David Lagnado, Joshua Tenenbaum. From counterfactual simulation to causal judgment,
36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.. 23-JUL-14, . : ,

09/29/2014 78.00 Brenden Lake, Ruslan Salakhutdinov, Joshua Tenenbaum. One-shot learning by inverting a compositional causalprocess,
Advances in Neural Information Processing Systems (NIPS). 05-DEC-13, . : ,

09/29/2014 79.00 Tomer Ullman, Andreas Stuhlmuller, Noah Goodman, Joshua Tenenbaum. Learning physics from dynamical scenes,
Thirty-Sixth Annual Conference of the Cognitive Science society.. 23-JUL-14, . : ,

09/29/2014 80.00 Tobias Gerstenberg, Tomer Ullman, Max Kleiman-Weiner, David Lagnado, Joshua Tenenbaum. Wins Above Replacement: Responsibility attributions as counterfactual replacements,
36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society. 23-AUG-14, . : ,

09/29/2014 82.00 James Lloyd, David Duvenaud, Roger Grosse, Joshua Tenenbaum, Zoubin Ghahramani. Automatic Construction and Natural-Language Description of Nonparametric Regression Models,
Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI-14). 27-JUL-14, . : ,

TOTAL: 80

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper
05/04/2010	15.00	Chris L. Baker, Rebecca Saxe, Joshua B. Tenenbaum. Action understanding as inverse planning, Cognition (01 2009)
05/04/2010	21.00	Sangho Park, Henry Kautz. Privacy-preserving Recognition of Activities in Daily Living from Multi-view Silhouettes and RFID-based Training, AAAI Fall 2008 Symposium on AI in Eldercare: New Solutions to Old Problems (01 2008)
05/04/2010	20.00	Jianqiang Shen, Erin Fitzhenry, Thomas G. Dietterich. Discovering Frequent Work Procedures From Resource Connections, Proceedings of the International Conference on Intelligent User Interfaces (IUI-2009) (02 2009)
05/04/2010	19.00	Jianqiang Shen, Thomas G. Dietterich. A Family of Large Margin Linear Classifiers and Its Application in Dynamic Environments, Proceedings of the SIAM International Conference on Data Mining 2009 (SDM-09) (01 2009)
05/04/2010	18.00	Jianqiang Shen, Jed Irvine, Xinlong Bao, Michael Goodman, Stephen Kolibaba, Anh Tran, Fredric Carl, Brenton Kirschner, Simone Stumpf, Thomas G. Dietterich. Detecting and Correcting User Activity Switches: Algorithms and Interfaces, Proceedings of the International Conference on Intelligent User Interfaces (IUI-2009) (02 2009)
05/04/2010	17.00	Victoria Keiser, Thomas G. Dietterich. Evaluating online text classification algorithms for email prediction in TaskTracer, Proceedings of (07 2009)
05/04/2010	16.00	Aniruddh Nath, Pedro Domingos. A Language for Relational Decision Theory, Proceedings of the Sixth International Workshop on Statistical Relational Learning (01 2009)
05/05/2010	22.00	Charles Kemp, Joshua B. Tenenbaum. The Discovery of Structural Forms, Proceedings of the National Academy of Sciences (08 2008)
05/05/2010	26.00	Charles Kemp, Noah D. Goodman, Joshua B. Tenenbaum. Theory Acquisition and the Language of Thought, (01 2008)
05/05/2010	25.00	Charles Kemp, Noah D. Goodman, Joshua B. Tenenbaum. Learning and using relational theories, Advances in Neural Information Processing Systems 20 (01 2008)
05/06/2010	28.00	Noah D. Goodman, Vikash K. Mansinghka, Daniel M. Roy, Keith Bonawitz, Joshua B. Tenenbaum. Church: A Language for Generative Models, (05 2008)
05/06/2010	31.00	Noah D. Goodman, Chris L. Baker, Joshua B. Tenenbaum. Cause and Intent: Social Reasoning in Causal Learning, (01 2009)
05/06/2010	30.00	Noah D. Goodman, Tomer D. Ullman, Joshua B. Tenenbaum. Learning a Theory of Causality, (01 2009)
05/06/2010	27.00	Chris L. Baker, Noah D. Goodman, Joshua B. Tenenbaum. Theory-based Social Goal Inference, (01 2008)

05/06/2010 29.00 Yarden Katz, Noah D. Goodman, Kristian Kersting, Charles Kemp, Joshua B. Tenenbaum. Modeling Semantic Cognition as Logical Dimensionality Reduction,
(01 2008)

05/21/2009 1.00 Rohit Kate, Raymond Mooney. Probabilistic Abduction using Markov Logic Networks,
IJCAI 2009 Workshop on Plan, Activity, and Intent Recognition (2009)

06/10/2009 4.00 Joseph Gonzalez, Yucheng Low, Carlos Guestrin. Residual Splash for Optimally Parallelizing Belief Propagation,
Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (2009)

06/10/2009 2.00 Jesse Davis, Pedro Domingos. Deep Transfer via Second-Order Markov Logic,
Proceedings of the Twenty-Sixth International Conference on Machine Learning (2009)

06/10/2009 5.00 Dafna Shahaf, Anton Chechetka, Carlos Guestrin. Learning Thin Junction Trees via Graph Cuts,
JMLR Workshop and Conference Proceedings Volume 5: Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS 2009) (2009)

06/10/2009 3.00 Stanley Kok, Pedro Domingos. Learning Markov Logic Network Structure via Hypergraph Lifting,
Proceedings of the Twenty-Sixth International Conference on Machine Learning (2009)

06/23/2009 6.00 Khalid El-Arini, Gaurav Veda, Dafna Shahaf, Carlos Guestrin. Turning Down the Noise in the Blogosphere,
The 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2009)

06/23/2009 10.00 Lilyana Mihalkova, Raymond Mooney. Learning to Disambiguate Search Queries from Short Sessions ,
Proceedings of the 2009 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (2009)

06/23/2009 9.00 Tuyen Huynh, Raymond Mooney. Max-Margin Weight Learning for Markov Logic Networks,
Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2009 (2009)

06/23/2009 8.00 Joseph Gonzalez, Yucheng Low, Carlos Guestrin, David OHallaron. Distributed Parallel Inference on Large Factor Graphs,
Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (2009)

06/23/2009 7.00 Hoifung Poon, Pedro Domingos. Unsupervised Semantic Parsing,
Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. (2009)

06/24/2009 12.00 Modayil, Rich Levinson, Craig Harman, David Halper, Henry Kautz. Integrating Sensing and cueing for More Effective Activity Reminders,
AAAI Fall 2008 Symposium on AI in Eldercare: New Solutions to Old Problems (2008)

06/24/2009 13.00 Parag Singla, Henry Kautz, Jiebo Luo, Andrew Gallagher. Discovery of Social Relationships in Consumer Photo Collections using Markov Logic,
3rd International Workshop on Semantic Learning and Applications in Multimedia Applications (2008)

06/25/2009 14.00 Pedro Domingos, Daniel Lowd. Markov Logic: An Interface Layer for Artificial Intelligence,
(01 2009)

09/03/2014 60.00 Daniel Lowd, Amirmohammad Rooshenas. The Libra Toolkit for Probabilistic Models,
Journal of Machine Learning Research (03 2014)

09/03/2014 61.00 Daniel Lowd, Brenton Lessley, Mino De Raj. Towards Adversarial Reasoning in Statistical Relational Domains,
AAAI-14 Workshop on Statistical Relational AI (04 2014)

09/15/2010 32.00 Jesse Davis, Pedro Domingos. Bottom-Up Learning of Markov Network Structure,
Proceeding of the 27th International Conference on Machine Learning (ICML 2010) (01 2010)

09/21/2010 34.00 Tuyen N. Huynh, Raymond J. Mooney. Online Max-Margin Weight Learning with Markov Logic Networks,
AAAI-10 Workshop on Statistical Relational AI (07 2010)

09/21/2010 33.00 Sindhu Raghavan, Raymond Mooney. Bayesian Abductive Logic Programs,
AAAI-10 Workshop on Statistical Relational AI (07 2010)

09/22/2010 35.00 Dafna Shahaf, Carlos Guestrin. Connecting the Dots Between News Articles,
The 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (07 2010)

09/22/2010 44.00 Jesse Davis, Pedro Domingos. Bottom-Up Learning of Markov Network Structure,
Proceeding of the 27th International Conference on Machine Learning (ICML 2010) (01 2010)

09/22/2010 43.00 Parag Singla, Aniruddh Nath, Pedro Domingos. Approximate Lifted Belief Propagation,
AAAI-10 Workshop on Statistical Relational AI (01 2010)

09/22/2010 42.00 Aniruddh Nath, Pedro Domingos. Efficient Lifting for Online Probabilistic Inference,
(01 2010)

09/22/2010 38.00 Yucheng Low, Joseph Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin, Joseph M. Hellerstein.
GraphLab: A New Parallel Framework for Machine Learning,
(09 2010)

09/22/2010 41.00 Aniruddh Nath, Pedro Domingos. Efficient Belief Propagation for Utility Maximization and Repeated Inference,
Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10) (01 2010)

09/22/2010 40.00 Hoifung Poon, Pedro Domingos. Unsupervised Ontology Induction from Text,
(07 2010)

09/22/2010 39.00 Stanley Kok, Pedro Domingos. Learning Markov Logic Networks Using Structural Motifs,
(01 2010)

09/22/2010 37.00 Joseph K. Bradley, Carlos Guestrin. Learning Tree Conditional Random Fields,
Proceeding of the 27th International Conference on Machine Learning (ICML 2010) (01 2010)

09/22/2010 36.00 Anton Chechetka, Carlos Guestrin. Focused Belief Propagation for Query-Specific Inference,
Thirteenth International Conference on Artificial Intelligence and Statistics (01 2010)

09/23/2010 48.00 Thomas G. Dietterich, Xinlong Bao, Victoria Keiser, Jianqiang Shen. Machine Learning Methods for High Level Cyber Situation Awareness,
Cyber Situational Awareness: Issues and Research (Advances in Information Security) (01 2010)

09/23/2014 75.00 Tianqi Chen, Sameer Singh, Carlos Guestrin. Gradient Boosting for Conditional Random Fields,
NIPS (submitted) 2014 (12 2014)

09/24/2010 50.00 Vibhav Gogate, Pedro Domingos. Exploiting Logical Structure in Lifted Probabilistic Inference,
AAAI-10 Workshop on Statistical Relational AI (01 2010)

09/24/2010 56.00 Tomer D. Ullman, Chris L. Baker, Owen Macindoe, Owain Evans, Noah D. Goodman, Joshua B. Tenenbaum. Help or Hinder: Bayesian Models of Social Goal Inference,
Advances in Neural Information Processing Systems 22 (NIPS 2009) (01 2010)

09/24/2010 55.00 Noah D. Goodman, Tomer D. Ullman, Joshua B. Tenenbaum. Learning a Theory of Causality,
Psychological Review (06 2010)

09/24/2010 54.00 Charles Kemp, Noah D. Goodman, Joshua B. Tenenbaum. Learning to Learn Causal Models,
Cognitive Science (06 2010)

09/24/2010 53.00 Leah Henderson, Noah D. Goodman, Joshua B. Tenenbaum, James F. Woodward. The Structure and Dynamics of Scientific Theories: A Hierarchical Bayesian Perspective,
Philosophy of Science (01 2010)

09/24/2010 52.00 Charles Kemp, Joshua B. Tenenbaum, Sourabh Niyogi, Thomas L. Griffiths. A probabilistic model of theory formation,
Cognition (01 2010)

09/24/2010	49.00	Vibhav Gogate, Pedro Domingos. Formula-Based Probabilistic Inference, Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (01 2010)
09/24/2010	51.00	Chloe Kiddon, Pedro Domingos. Leveraging Ontologies for Lifted Probabilistic Inference and Learning, AAAI-10 Workshop on Statistical Relational AI (01 2010)
09/26/2010	63.00	Jerry R. Hobbs, Rutu Mulkar-Mehta. Toward a Formal Theory of Information Structure, Evolution of Semantic Systems (01 2010)
09/26/2010	64.00	Andrew S. Gordon, Jerry R. Hobbs, Michael T. Cox. Anthropomorphic Self-Models for Metareasoning Agents, (01 2010)
09/26/2010	65.00	Jerry R. Hobbs. Questions in Decision-Making Dialogues, Questions: Formal, Functional and Interactional Perspectives, Cambridge University Press Series on Language, Culture, and Cognition (01 2010)
09/26/2010	66.00	Adam Sadilek, Henry Kautz. Recognizing Multi-Agent Activities from GPS Data, Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10) (01 2010)
09/26/2010	67.00	Adam Sadilek, Henry Kautz. Modeling and Reasoning about Success, Failure and Intent of Multi-Agent Activities, UbiComp 2010 Workshop on Mobile Context-Awareness (09 2010)
09/26/2010	68.00	Alan L. Liu, Harlan Hile, Gaetano Borriello, Pat A. Brown, Mark Harniss, Henry Kautz, Kurt Johnson. Customizing Directions in an Automated Wayfinding System for Individuals with Cognitive Impairment, Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009) (10 2009)
09/26/2010	57.00	Leon Bergen, Owain R. Evans, Joshua B. Tenenbaum. Learning Structured Preferences, Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (01 2010)
09/26/2010	58.00	Tomer D. Ullman, Noah D. Goodman, Joshua B. Tenenbaum. Theory Acquisition as Stochastic Search, Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (01 2010)
09/26/2010	59.00	Jerry R. Hobbs. Clause-Internal Coherence, Constraints in Discourse 2 (01 2010)
09/26/2010	60.00	Jerry R. Hobbs, Ellen Riloff. Information Extraction, Handbook of Natural Language Processing, Second Edition (01 2010)
09/26/2010	61.00	Jerry R. Hobbs, Andrew Gordon. Goals in a Formal Theory of Commonsense Psychology, Proceedings of the Sixth Formal Ontologies in Information Systems Conference (FOIS-2010) (05 2010)
09/26/2010	62.00	Jerry R. Hobbs. Word Meaning and World Knowledge, Handbook of Semantics (09 2010)
09/27/2010	69.00	Hoifung Poon, Pedro Domingos. Machine Reading: A "Killer App" for Statistical Relational AI , AAAI-10 Workshop on Statistical Relational AI (07 2010)
10/20/2011	70.00	Dror Baron, Alexander Ihler, Harel Avissar, Danny Bickson, Danny Dolev. Fault identification via nonparametric belief propagation, IEEE Tran. on Signal Processing vol. 59, no. 6, pp. 2602-2603, June 2011 (10 2011)

TOTAL: **67**

Number of Manuscripts:

Books

Received	Book
08/24/2012 93.00	Adam Sadilek, Henry Kautz. "Modeling Success, Failure, and Intent ofMulti-Agent Activities Under Severe Noise" in "Mobile Context Awareness," Tom Lovett and Eamonn O'Neill (Editors), New York: Springer, (04 2012)
08/25/2014 47.00	Sindhu Raghavan, Parag Singla, Raymond J. Mooney. Plan Recognition using Statistical Relational Models, United States of America: Morgan Kaufmann, (03 2014)

TOTAL: **2**

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Pedro Domingos won the ACM SIGKDD Innovation Award, the highest honor in the field of knowledge discovery and data mining.

Pedro Domingos' graduate student Robert Gens won the Google Deep Learning Fellowship.

Pedro Domingos: Invited Speaker, Twentieth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, 2014.

Pedro Domingos: Invited Tutorial, Thirtieth Conference on Uncertainty in Artificial Intelligence, with Daniel Lowd, Quebec City, Canada, 2014.

Pedro Domingos: Instructor, Twenty-Fifth Machine Learning Summer School, Beijing, China, 2014.

Pedro Domingos: Invited Speaker, Duke University, Durham, NC, 2014.

Pedro Domingos: Invited Speaker, Allen Institute for Artificial Intelligence, Seattle, WA, 2014.

Pedro Domingos: Invited Speaker, Microsoft Research, Redmond, WA, 2014.

Pedro Domingos: Invited Speaker, Second International Conference on Learning Representations, Banff, Canada, 2014.

Pedro Domingos: Invited Speaker, First IKDD Conference on Data Sciences, New Delhi, India, 2014.

Pedro Domingos: Invited Speaker, International Conference on Machine Learning and Applications, Miami Beach, FL, 2013.

Pedro Domingos: Invited Speaker, Eighth Workshop on Graph-Based Methods for Natural Language Processing, Seattle, WA, 2013.

Vibhav Gogate: Invited Speaker, "Fast, Lifted Sampling Algorithms" at the 2014 AAAI Workshop on Statistical Relational Artificial Intelligence.

Daniel Lowd: Invited Speaker, "Using Dependency Networks To Learn Markov Networks", University of Wisconsin-Madison. 2013/10.

Daniel Lowd: Invited Speaker, "Learning Relational Classifiers for Adversarial Domains" University of Indiana-Bloomington. 2013/10.

Daniel Lowd: Invited Speaker, "Learning Relational Classifiers for Adversarial Domains" SRI International. 2013/10.

Tom Dietterich: Keynote Speaker, International Conference on Artificial General Intelligence (AGI-2013), Beijing, July 30, 2013. "Reflections on CALO: General Intelligence for the Desktop"

Tom Dietterich: Invited Talk, First International Conference on Reinforcement Learning and Decision Making (RLDM-2013), Princeton University, October 27, 2013, "Simulator-defined MDPs in Ecosystem Management"

Tom Dietterich: Distinguished Speaker, Columbia University Data Science Institute, March 13, 2014, "Challenges for Machine Learning in Computational Sustainability"

Tom Dietterich: Distinguished Speaker, University of San Francisco, April 11, 2014, "Challenges for Machine Learning in Computational Sustainability"

Tom Dietterich: Keynote Speaker, Computational Modeling Showcase, Oberlin College, Oberlin, Ohio, May 8, 2014, "Modeling bird migration by combining weather radar and citizen science data"

Tom Dietterich: Invited Speaker, Signatures Lecture Series, Pacific Northwest National Labs, June 9, 2014, "Advances in Anomaly Detection"

Tom Dietterich: Invited Lecture, Machine Learning Summer School, Beijing, China, June 16, 2014, "Introduction to Machine Learning"

Tom Dietterich: Invited Speaker, Zhejiang Sci-Tech University, Hangzhou, China, June 17, 2014, "Computational Ecology and Ecosystem Management"

Tom Dietterich: Invited Speaker, China National Rice Research Institute, Hangzhou, China, June 18, 2014, “Computer Vision for Insect Population Counting: Project BugID”

Tom Dietterich: Elected President, Association for the Advancement of Artificial Intelligence

Joshua Tenenbaum: Elected Fellow of the Cognitive Science Society, 2013

Joshua Tenenbaum: Invited keynote speaker, CVPR 2014 workshop, “Vision meets Cognition”, June 2014

Joshua Tenenbaum: Invited speaker, Annual Meeting of the Cognitive Science Society, Plenary Symposium on Compiutatinal Models of Moral Cognition, July 2014

Joshua Tenenbaum: Invited speaker, University of Massachusetts, Amherst, Computational Social Science Program, April 2014

Joshua Tenenbaum: Invited speaker, Columbia University, Theoretical Neuroscience Colloquium, April 2014

Joshua Tenenbaum: Invited speaker, Carnegie Mellon University, Machine Learning Department Distinguished Speaker Series, March 2014

Joshua Tenenbaum: Invited speaker, Carnegie Mellon University, Center for the Neural Basis of Cognition Colloquium, March 2014

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Tianqi Chen	0.11	
Robert Gens	0.24	
William Webb	0.04	
Nicholas Wilson	0.50	
Subhashini Venugopalan	0.07	
Deepak Venugopal	0.75	
Tahrima Rahman	0.75	
Mohamad Ali Torkamani	1.00	
Amirmohammad Rooshenas	0.61	
Marco Correia Ribeiro	0.22	
Jesse Hostetler	0.62	
Tomer Ullman	0.33	
FTE Equivalent:	**5.24**	
Total Number:	**12**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Mathias Niepert	1.00
Tobias Gerstenberg	0.33
FTE Equivalent:	**1.33**
Total Number:	**2**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Pedro Domingos	0.32	
Vibhav Gogate	0.08	
Joshua Tenenbaum	0.22	
FTE Equivalent:	**0.62**	
Total Number:	**3**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME	
Nicholas Wilson	
Haijie Gu	
Total Number:	**2**

Names of personnel receiving PHDs

NAME
Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
Patrick Allen	0.08
Andrei Stabrovski	0.20
Stacy Miller	0.13
Peter Battaglia	0.20
Jennifer White	0.01
FTE Equivalent:	**0.62**
Total Number:	**5**

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

INTRODUCTION AND PROGRESS OVERVIEW

The goal of this MURI project is to develop a unified approach to abductive inference, which combines the capabilities of logic and probability. Abduction is inference to the best explanation. Consider the situation of a military commander. He needs to make sense of a bewildering array of information: sensors carried by troops on patrol, sensors placed along roads, video from surveillance cameras, video and other feeds from unmanned vehicles (aerial and ground), remote sensing streams, intercepted communications, mission reports, intelligence reports, news stories, etc. Critical decisions depend on the correct interpretation of this data: Where to send troops? Should a convoy be rerouted? Is an attack imminent? Is the behavior of a group of people suspicious? Bridging the gulf between the mass of low-level sensor information and the necessary high-level understanding of a situation requires abductive inference. Humans are experts at this, but they can only handle so much information at a time (and also have well-documented biases and blind spots). Making the most of the available information requires automated inference that continuously supports and complements the commander. However, the scale, uncertainty and complexity of the task puts it beyond the reach of current AI systems.

An effective abductive inference system has to interpret evidence, construct explanations, discard irrelevant data (but revisit it if becomes relevant), integrate information from many different sources, formulate and test hypotheses, suggest alternative courses of action, and help direct the collection of further data. It must be able to reason at every scale, from interpreting the moment-by-moment actions of the enemy and the population it is embedded into detecting plans whose steps may be far apart in time (e.g., scoping a target location, procuring materials for a bomb, assembling it, deploying it, etc.), to understanding the motivations and thought processes of the enemy.
The goal of the research proposed here is to make such a system a reality. We are developing a well-founded approach to abduction based on the solid foundations of first-order logic and probability. We are using Markov logic (Domingos and Lowd, 2009), which unifies first-order logic and probability, as the common representation language.
The research program we proposed has four major components: foundations, inference, learning, and activity and plan recognition. We completed the proposed research on foundations in the first three years. In the following, we will describe our accomplishments and progress made in the past reporting period on the latter three components, starting with a brief overview.

INFERENCE

This component of the project seeks to build scalable, next-generation inference systems for Markov logic but also particular variations of Markov logic that render inference and/or learning more tractable.

We have made progress in the following areas:

Fast Evidence Processing in Markov Logic: Evidence breaks symmetries and lifted inference algorithms, state of the art inference methods for statistical relational models, end up grounding the MLN. To solve these problems, we have developed a general method to achieve scalable inference in MLNs, allowing for arbitrary structures with arbitrary evidence. We reduced the size of the inference problem of Markov Logic Networks by clustering together similar evidence atoms, and replacing all atoms in each cluster by their cluster center. To formulate the clustering problem, we have developed several similarity measures. We performed experiments on several different benchmark MLNs utilizing various clustering and inference algorithms. Our experiments clearly show the generality and scalability of our approach.
Lifted MAP Inference for Markov Logic: We improved existing lifted MAP inference results further by showing that if non-shared MLNs contain no self joins, namely every atom appears at most once in each of its formulas, then all variables in the corresponding Markov network need only be bi-valued. Our approach is quite general and can be easily applied to arbitrary MLNs by simply grounding all of its shared terms. The key feature of our approach is that because we reduce lifted inference to propositional inference, we can use any propositional MAP inference algorithm for performing lifted MAP inference, without actually lifting the propositional algorithm.
Loopy Belief Propagation in presence of determinism: We proposed a new method for improving the performance of loopy belief propagation in the presence of logical constraints and determinism. The key idea in our method is nding a reparameterization of the graphical model such that LBP, when run on the reparameterization, is likely to have better convergence properties than LBP on the original graphical model. We proposed several new schemes for nding such reparameterizations, all of which leverage unique properties of zeros as well as research on LBP convergence done over the last decade. Our experimental evaluation on a variety of PGMs clearly demonstrates the promise of our method -- it often yields accuracy and convergence time improvements of an order of magnitude or more over standard LBP.
Probabilistic Inference for Hybrid MLNs: We have continued to develop an efficient algorithm for continuous, nonconvex optimization, which we call RDIS. RDIS allows us to efficiently perform MAP inference in continuous domains and to accurately fit models with continuous parameters to data. RDIS easily supports discrete variables as well as continuous, enabling it to play an important role in multi-modal and multi-scale domains. The key to RDIS is to identify and exploit local structure in the objective function by dynamically identifying a subset of variables that, once optimized, decomposes the remaining variables into approximately independent subsets. These can then be separately optimized, and we do so recursively in order to exploit multiple levels of local structure, dynamically finding within each a subset that further decomposes it, etc.
Tractable Probabilistic Knowledge Bases (TPKB): Tractable Markov Logic was originally designed to be used with Probabilistic

Theorem Proving (PTP) as an inference algorithm, but PTP was much more complicated than necessary for the restricted case, which clouded intuition and made formal guarantees difficult. Last year, we worked on improvements of TML that feature existence uncertainty and an object oriented syntax. This year, we further improved TPKBs and learned a large-scale TPKB from a variety of data sources, a longstanding goal of AI research. Existing approaches either ignore the uncertainty inherent to knowledge extracted from text, the web, and other sources, or lack a consistent probabilistic semantics with tractable inference. TPKBs consist of a hierarchy of classes of objects and a hierarchy of classes of object pairs such that attributes and relations are independent conditioned on those classes. These characteristics facilitate both tractable probabilistic reasoning and tractable maximum-likelihood parameter learning. TPKBs feature a rich query language that allows one to express and infer complex relationships between classes, relations, objects, and their attributes. The queries are translated to sequences of operations in a relational database facilitating query execution times in the sub-second range. We demonstrate the power of TPKBs by leveraging large data sets extracted from Wikipedia to learn their structure and parameters. The resulting TPKB models a distribution over millions of objects and billions of parameters. We apply the TPKB to entity resolution and object linking problems and show that the TPKB can accurately align large knowledge bases and integrate triples from open IE projects.
Alchemy 2.0 and Alchemy Lite: We continued to work on Alchemy (our open source software package for learning and inference in Markov Logic). The main highlight of the newer version is access to highly scalable lifted inference algorithms, lifted sampling approaches developed in this project over the past two years. We also continued to work on Alchemy Lite, an open-source software package for inference in Tractable Markov Logic (TML), the first tractable first-order probabilistic logic. For a list of companies and people that have used Alchemy, we refer the reader to the end of the report.

LEARNING

The goal of learning algorithms is to acquire knowledge from data autonomously, without human intervention. In the past year, we have made progress on the following fronts.

Structure Learning for Sum-Product Networks: We developed a new SPN structure learning algorithm, called ID-SPN, for learning SPNs with both indirect and direct variable interactions. ID-SPN performs a top-down clustering of instances and variables, similar to our previous method, but with tractable graphical models at the leaves instead of univariate distributions. These leaf distributions are learned using ACMN, a state-of-the-art method we developed for learning arithmetic circuits (Lowd & Rooshenas, 2013). ID-SPN uses the likelihood of a validation set to determine the proper depth of each branch. In experiments on a standard set of benchmark datasets, ID-SPN is more accurate than the previous state-of-the-art on 20 out of 20 datasets. ID-SPN also obtains better likelihoods than intractable Bayesian networks on 13 out of 20 datasets, suggesting that tractable models can be as accurate as intractable ones.
Sum-Product Networks for Vision: We investigated the use of SPNs for fast object recognition and three-dimensional pose estimation. We continued to investigate the use of Sum-Product Networks for object recognition and pose estimation. As a step toward this goal, we developed the Deep Symmetry Network (DSN), a neural network that can model invariance to richer transformations.
Deep Symmetry Networks: We developed deep symmetry networks (symnets), a generalization of convnets that forms feature maps over arbitrary symmetry groups. Symnets use kernel-based interpolation to tractably tie parameters and pool over symmetry spaces of any dimension. They also use a Lucas-Kanade optimization to warp features to feature maps. Experiments on the NORB and MNIST-rot datasets show that symnets over the affine group greatly reduce sample complexity relative to convnets by better capturing the symmetries in the data.
Symmetry-based Semantic Parsing: We have developed a new notion of semantics based on symmetry group theory. Our new approach allows us to develop a semantic parser that avoids the challenges of having to choose a formal meaning representation or having to gather large amounts of labeled training data; our parser also represents semantics in a way that is easily integrated into a Tractable Probabilistic Knowledge Base over which abductive inference will be performed.
Exchangeable Variable Models: Conditional independence is a crucial notion that facilitates efficient inference and parameter learning in probabilistic models. Its logical and algorithmic properties as well as its graphical representations have led to the advent of graphical models as a discipline within artificial intelligence. The notion of finite (partial) exchangeability (Diaconis & Freedman, 1980a), on the other hand, has not yet been explored as a basic building block for tractable probabilistic models. A sequence of random variables is exchangeable if its distribution is invariant under variable permutations. Similar to conditional independence, partial exchangeability, a generalization of exchangeability, can reduce the complexity of parameter learning and is a concept that facilitates high tree-width graphical models with tractable inference.
Relational Sum-Product Networks: Building on our earlier work on Relational Sum-Product Networks, we developed LearnRSPN, the first algorithm for learning high-treewidth tractable statistical relational models. LearnRSPN is a recursive top-down structure learning algorithm for RSPNs, based on the LearnSPN (Gens and Domingos 2013) algorithm for propositional SPN learning. In our empirical evaluation, the RSPN learning algorithm outperforms Markov Logic Networks (Richardson and Domingos 2006) in both running time and predictive accuracy.
Tractable Probabilistic Programs: We began to develop Tractable Probabilistic Programs (TPP), a statistical relational representation that captures a probability distribution over programs drawn from a context-free grammar. Although TPP was motivated by the problem of automated software debugging, we are investigating its applicability to other problem domains that make use of probabilistic grammars, such as natural language processing and image understanding.

We proposed a novel model for Biomedical event extraction based on MLNs that leverages the power of support vector machines (SVMs) to handle high-dimensional features. Specifically, we learned SVM models using rich linguistic features for trigger and argument detection and type labeling; designed an MLN composed of soft formulas (each of which encodes a soft constraint whose associated weight indicates how important it is to satisfy the constraint) and hard formulas (constraints that always need to be satisfied, thus having a weight to capture the relational dependencies between triggers and arguments; and encoded the SVM output as prior knowledge in the MLN in the form of soft formulas, whose weights are computed using the confidence values generated by the SVMs. This formulation naturally allows SVMs and MLNs to complement each other's strengths and weaknesses: learning in a large and sparse feature space is much easier with SVMs than with MLNs, whereas modeling relational dependencies is much easier with MLNs than with SVMs.

BLPs for Textual Inference: During the final year of the project, the research at the University of Texas at Austin has focused on applying and adapting the statistical-relational AI techniques we developed under the MURI project to the problem of abductive inference for natural-language text understanding, which we are investigating as part of DARPA's Deep Exploratory and Filtering of Text (DEFT) program. We used the Bayesian Logic Programming (BLP) methods we developed to learn knowledge-bases for making probabilistic inferences from information automatically extracted from text.

Distributional Markov Logic Semantics: We applied some of the Markov logic methods we developed to construct a general formal semantics for natural language that integrates traditional logical form produced by a broad-coverage parser with probabilistic rules extracted from a vector-space distributional semantics automatically constructed from large corpora. These techniques were evaluated on the Knowledge Base Population (KBP) formal evaluation conducted by NIST and on standardized datasets for Recognizing Textual Entailment (RTE) and Semantic Textual Similarity (STS).

Robust Structured Prediction through Regularization: In previous work, we developed max-margin learning methods for collective classification that are robust to adversarial manipulation of object features (Torkamani & Lowd, 2013). However, these methods were restricted to associative Markov networks and could not handle more complex scenarios, such as adversaries that manipulate link structure. We developed a new strategy for learning robust Markov networks or structural SVMs by showing that robustness to perturbations of the features is equivalent to regularization. Specifically, when perturbations are constrained by a norm, the equivalent regularizer is given by the dual norm. When perturbations are constrained by a polyhedron, the equivalent regularizer is a linear function in a transformed space. In experiments, we demonstrate that this regularization strategy leads to improved generalization on a collective classification problem with a lot of concept drift.

Large-Scale Machine Learning: We focused on the long term goal of building and improving our GraphLab large scale parallel machine learning framework (http://graphlab.org). We extended the GraphLab framework to the substantially more challenging distributed setting while preserving strong data consistency guarantees. Two strong areas of focus have been in graphical models and parallel learning. To address these problems in a more accurate fashion, we've developed a gradient boosting algorithm for tree-shaped conditional random fields (CRF). Conditional random fields are an important class of models for accurate structured prediction, but effective design of the feature functions is a major challenge when applying CRF models to real world data. Gradient boosting, which can induce and select functions, is a natural candidate solution for the problem. However, it is non-trivial to derive gradient boosting algorithms for CRFs, due to the dense Hessian matrices introduced by variable dependencies. We address this challenge by deriving a Markov Chain mixing rate bound to quantify the dependencies, and introduce a gradient boosting algorithm that iteratively optimizes an adaptive upper bound of the objective function. The resulting algorithm induces and selects features for CRFs via functional space optimization, with provable convergence guarantees. Experimental results on three real world datasets demonstrate that the mixing rate based upper bound is effective for training CRFs with non-linear potentials.

ACTIVITY AND PLAN RECOGNITION

Activity and plan recognition is a classic problem of abductive inference. The goal is to infer the goals and plans of one or more agents from noisy and fragmentary observations of their behaviour. During the reporting period, we made significant progress on both applied and fundamental work on plan and activity recognition:

Kitchen activities: We implemented and evaluated a Markov-logic based plan recognition system for kitchen activities, where observations came from video and natural language narration.

Disease Prediction with Social Media Data: We extended our work on activity and state recognition from social media data. We showed that we could model disease transmission at a global scale using posts from airline travellers, and could identify possible sources of food poisoning by identifying restaurant meal events and sickness events from Twitter posts.

More Efficient Modal Markov Logic: We developed complexity results on a "less intractable" subset of multi-agent Markov Logic.

Plan Recognition with Monte Carlo Tree Search: We developed the mathematical foundations for state abstraction in MCTS. We showed that state abstraction in MCTS is much easier to achieve. We proved accuracy bounds for a certain form of state abstraction, state aggregation in ExpectiMax trees, and we showed that these state abstractions preserve optimality in search trees. This in turn permitted us to prove correctness of state aggregation abstractions for two MCTS methods: Sparse Sampling and UCT. These results are very general and show excellent performance improvements in several benchmark problems. Future work will focus on automatically learning these state abstractions.

BACKGROUND: MARKOV LOGIC

Markov logic (Domingos and Lowd, 2009) provides the foundation for our research into abductive inference. It is unique in its simplicity and generality, and in the range, scalability and sophistication of its algorithms, which are publicly available in the open-source Alchemy package. Markov logic attaches weights to formulas in first-order logic. A first-order knowledge base can be seen as a set of hard constraints on the set of possible worlds: if a world violates even one formula, it has zero probability. The basic idea in Markov logic is to soften these constraints: when a world violates one formula in the knowledge base it is less probable, but not impossible. The fewer formulas a world violates, the more probable it is. A formula's associated weight reflects how strong a constraint it is: the higher the weight, the greater the difference in log probability between a world that satisfies the formula and one that does not, other things being equal. We call a set of weighted first-order formulas a Markov logic network (MLN). Semantically, we view the formulas as templates for constructing Markov networks (the undirected counterpart of Bayes nets). Given different sets of constants, an MLN will produce different networks, and these may be of widely varying size, but all will have certain regularities in structure and parameters (e.g., all groundings of the same formula will have the same weight). Markov logic has first-order logic and most discrete statistical models used in AI as special cases. Alchemy currently includes algorithms for inferring the most probable explanation of evidence, computing marginal and conditional probabilities, learning formula weights from data (generatively and discriminatively), and learning and/or revising formulas.

BACKGROUND: TRACTABLE MARKOV LOGIC (TML)

We have to date made much progress on combining first-order logic with probability, starting with Markov Logic and more recently with Probabilistic Theorem Proving (PTP) (Gogate and Domingos, 2011). However, both of these approaches suffer from the fact that even simple models created with them are often intractable, which essentially precludes their widespread use. Sum-Product Networks (SPNs) (Poon and Domingos, 2011) have guaranteed tractability, but at the expense of greatly reduced expressiveness, specifically, they are not first-order and can be difficult to interpret. Tractability and expressiveness seem to be fundamentally opposed and one might think that having useful first-order features would necessarily allow for intractable models, however, we have devised a new language that we believe combines substantial first-order representational richness with guaranteed tractability. This language, which we call Tractable Markov Logic (TML) is a subset of Markov Logic whose structure was informed by both SPNs and PTP. TML can represent objects and relations in a first-order fashion and is structured according to ontology-like class and part hierarchies. These restrictions are strong enough to allow for an exact inference algorithm that is linear in the number of objects times the number of rules in the knowledge base. However, they are also weak enough that TML can compactly represent essentially all widely-used tractable models, including junction trees, probabilistic context-free grammars and SPNs. Additionally, TML permits probabilistic versions of inheritance hierarchies and default reasoning. These results are described in our paper (Domingos and Webb, 2012), implemented in preliminary form as a software package called Alchemy Lite, will be presented soon.

PROGRESS ON INFERENCE

LOOPY BELIEF PROPAGATION IN THE PRESENCE OF LOGICAL DEPENDENCIES

It is well known that loopy Belief propagation (LBP), perhaps the most widely used and the most researched inference algorithm, performs poorly on probabilistic graphical models (PGMs) with determinism and logical dependencies. This is problematic because many probabilistic programs contain large amount of determinism and logical constraints. Therefore, in this work, we proposed a new method for remedying this problem. The key idea in our method is finding a reparameterization of the graphical model such that LBP, when run on the reparameterization, is likely to have better convergence properties than LBP on the original graphical model. We proposed several new schemes for finding such reparameterizations, all of which leverage unique properties of zeros as well as research on LBP convergence done over the last decade. Our experimental evaluation on a variety of PGMs clearly demonstrates the promise of our method -- it often yields accuracy and convergence time improvements of an order of magnitude or more over LBP. This work was published at the 2014 AISTATS conference.

FAST EVIDENCE PROCESSING IN MARKOV LOGIC NETWORKS

Markov Logic Networks (MLNs) unify first order logic and probabilistic graphical models. However, due to the rich representational power of MLNs, inference in these models is extremely challenging. The standard graphical model inference algorithms operate on the ground model and do not scale well as the number of objects gets larger. On the other hand, lifted inference algorithms which perform inference at the first-order level offer the desired scalability. However, the conditions under which the MLN can be correctly processed at the lifted level are often very restrictive. A worse problem is that, evidence breaks symmetries and lifted inference algorithms end up grounding the MLN. To solve these problems, we have developed a general

method to achieve scalable inference in MLNs, allowing for arbitrary structures with arbitrary evidence.
Our approach works is quite straight-forward. Given a MLN and a large set of evidence atoms, we reduce the size of the inference problem by clustering together similar evidence atoms, and replacing all atoms in each cluster by their cluster center. To formulate the clustering problem, we have developed several similarity measures. We performed experiments on several different benchmark MLNs utilizing various clustering and inference algorithms. Our experiments clearly show the generality and scalability of our approach. This work will appear in the 2014 ECML conference.

LIFTED MAP INFERENCE

We developed a new approach for approximate MAP inference in Markov Logic Networks (MLNs) (this work was published at the 2014 AISTATS conference). Our approach is based on the following key result that we proved: if an MLN has no shared terms then MAP inference over it can be reduced to MAP inference over a Markov network having the following properties: (i) the number of random variables in the Markov network is equal to the number of first-order atoms in the MLN; and (ii) the domain size of each variable in the Markov network is equal to the number of groundings of the corresponding first-order atom. This represents exponential complexity reductions over ground MAP inference. We improved this result further by showing that if non-shared MLNs contain no self joins, namely every atom appears at most once in each of its formulas, then all variables in the corresponding Markov network need only be bi-valued. .
Our approach is quite general and can be easily applied to arbitrary MLNs by simply grounding all of its shared terms. The key feature of our approach is that because we reduce lifted inference to propositional inference, we can use any propositional MAP inference algorithm for performing lifted MAP inference, without actually lifting the propositional algorithm. Within our approach, we experimented with two propositional MAP inference algorithms: Gurobi (an Integer Linear Programming solver) and MaxWalkSAT. Our experiments on several benchmark MLNs clearly demonstrated the superiority of our approach over the ground inference in terms of both the scalability and the solution quality.

PROBABILISTIC INFERENCE FOR HYBRID MLNS

General abductive inference requires the ability to reason about both discrete and continuous information. Accordingly, we have continued to develop an efficient algorithm for continuous, nonconvex optimization, which we call RDIS. RDIS allows us to efficiently perform MAP inference in continuous domains and to accurately fit models with continuous parameters to data. RDIS easily supports discrete variables as well as continuous, enabling it to play an important role in multi-modal and multi-scale domains. The key to RDIS is to identify and exploit local structure in the objective function by dynamically identifying a subset of variables that, once optimized, decomposes the remaining variables into approximately independent subsets. These can then be separately optimized, and we do so recursively in order to exploit multiple levels of local structure, dynamically finding within each a subset that further decomposes it, etc. RDIS is similar in structure to existing combinatorial algorithms and we accordingly use ideas from these and dynamically choose variables using hypergraph partitioning. This ensures that decomposition is achieved at each level, if at all possible. For value selection, RDIS employs ideas from continuous optimization in the form of a local optimization subroutine such as gradient descent or quasi-Newton to ensure that it can find the global optimum without needing to explore the infinitely-many values in the continuous domain. We've evaluated RDIS both analytically and empirically. Analytically, we show that RDIS finds the global minimum in exponentially less time than standard methods for a class of nonconvex functions that exhibit local structure. Empirically, tests on highly multi-modal functions, structure from motion, and protein folding demonstrate that our algorithm consistently finds significantly better minima than these same standard methods in challenging optimization and inference tasks.

TRACTABLE PROBABILISTIC KNOWLEDGE BASES (TPKB)

Tractable Markov Logic was originally designed to be used with Probabilistic Theorem Proving (PTP) as an inference algorithm, but PTP was much more complicated than necessary for the restricted case, which clouded intuition and made formal guarantees difficult. Last year, we worked on improvements of TML that feature existence uncertainty and an object oriented syntax. This year, we further improved TPKBs and learned a large-scale TPKB from a variety of data sources, a longstanding goal of AI research. While our large-scale probabilistic knowledge base is not the first large scale representation of such data, existing approaches either ignore the uncertainty inherent to knowledge extracted from text, the web, and other sources, or lack a consistent probabilistic semantics with tractable inference. TPKBs consist of a hierarchy of classes of objects and a hierarchy of classes of object pairs such that attributes and relations are independent conditioned on those classes. These characteristics facilitate both tractable probabilistic reasoning and tractable maximum-likelihood parameter learning. TPKBs feature a rich query language that allows one to express and infer complex relationships between classes, relations, objects, and their attributes. For instance, our query language features unions of existentially quantified conjunctive queries. These queries are translated to sequences of operations in a relational database facilitating query execution times in the sub-second range. We demonstrate the power of TPKBs by leveraging large data sets extracted from Wikipedia to learn their structure and parameters. The resulting TPKB models a distribution over millions of objects and billions of parameters. We apply the TPKB to entity resolution and object linking problems and show that the TPKB can accurately align large knowledge bases and integrate triples from open IE projects.

ALCHEMY 2.0 & ALCHEMY LITE

We continued to work on Alchemy (our open source software package for learning and inference in Markov Logic). The main highlight of Alchemy 2.0 is access to more scalable lifted inference algorithms such as lifted sampling approaches. We have also continued to work on algorithms that allow us to apply lifted inference algorithms to models that are usually not liftable. We also continued to develop Alchemy Lite, an open-source software package for inference in Tractable Markov Logic (TML), the first tractable first-order probabilistic logic. TML strikes a good balance between expressiveness and tractability, subsuming essentially all tractable models, including many high-treewidth ones. The software allows users to build intuitive models in an object-oriented style while guaranteeing that inference will be efficient without resorting to approximation or ad hoc performance hacks. Alchemy Lite allows for fast, exact inference for models formulated in terms of TML, as well as the ability to update models with new information. Further improvements to the inference implementation to allow for tasks such as entity resolution and parsing have also been under development.
For an extensive list of industry applications we refer the reader to the end of the report.

PROGRESS ON LEARNING

IMPROVED STRUCTURE LEARNING FOR SUM-PRODUCT NETWORKS

Learning the structure of SPNs is important for applying them to domains where the structure of the relationships among the variables is not known in advance. Our previous state-of-the-art algorithm (Gens & Domingos, 2013) for learning SPN structure performed top-down clustering of training instances and variables to create sum and product nodes, respectively. Variable interactions are represented indirectly through sum nodes, which act as latent variables. In contrast, most algorithms for learning graphical models represent interactions directly through conditional probability distributions or potential functions, not through latent variables.
We developed a new SPN structure learning algorithm, called ID-SPN, for learning SPNs with both indirect and direct variable interactions. ID-SPN performs a top-down clustering of instances and variables, similar to our previous method, but with tractable graphical models at
the leaves instead of univariate distributions. These leaf distributions are learned using ACMN, a state-of-the-art method we developed for learning arithmetic circuits (Lowd & Rooshenas, 2013). ID-SPN uses the likelihood of a validation set to determine the proper depth of each branch. In experiments on a standard set of benchmark datasets, ID-SPN is more accurate than the previous state-of-the-art on 20 out of 20 datasets. ID-SPN also obtains better likelihoods than intractable Bayesian networks on 13 out of 20 datasets, suggesting that tractable models can be as accurate as intractable ones.

DEEP SYMMETRY NETWORKS

The chief difficulty in object recognition is that objects' classes are obscured by a large number of extraneous sources of variability, such as pose and part deformation. These sources of variation can be represented by symmetry groups, sets of composable transformations that preserve object identity. Our goal is to combine the rich tractable inference of Sum-Product Networks with the generalization of symmetry groups. Deep Symmetry Networks show the benefit of extending neural networks to richer symmetry spaces. Convolutional neural networks (convnets) achieve a degree of translational invariance by computing feature maps over the translation group, but cannot handle other groups. As a result, these groups' effects have to be approximated by small translations, which often requires augmenting datasets and leads to high sample complexity. We developed deep symmetry networks (symnets), a generalization of convnets that forms feature maps over arbitrary symmetry groups. Symnets use kernel-based interpolation to tractably tie parameters and pool over symmetry spaces of any dimension. They also use a Lucas-Kanade optimization to warp features to feature maps. These techniques sidestep the exponential computational burden of convolving a feature in high dimensions. Like convnets, symnets are trained with backpropagation. The composition of feature transformations through the layers of a symnet provides a new approach to deep learning. Our preliminary experiments on the NORB and MNIST-rot datasets show that symnets over the affine group greatly reduce sample complexity relative to convnets by better capturing the symmetries in the data.

SYMMETRY-BASED SEMANTIC PARSING

An abductive inference system should be capable of interpreting all kinds of information. Many times, information will come into the system in the form of unstructured text. Typically, the strategy for dealing with unstructured text is use a semantic parser to map the text to its formal meaning representation in some first-order logic language. However, there is little consensus about the best meaning representation to choose and finding enough labeled training data to train such a semantic parser is difficult and costly. We have proposed a new notion of semantics that avoids these challenges, and, as added benefit, represents meaning in a way that is more easily integrated into a Tractable Probabilistic Knowledge Base (TPKB). We utilize insights from symmetry group theory, which studies the formal properties of symmetry groups, which are groups of transformations under which key properties of a structure are preserved. We introduce the concept of a semantic symmetry group, which contains syntactic operations which when applied to a sentence preserve its meaning. A semantic symmetry group partitions the set of all sentences into sets, called orbits, of sentences with the same meaning. The orbit that a sentence is a member of implicitly defines its meaning. Since natural language frequently contains ambiguities, we utilize a probabilistic approach to semantic symmetry and orbit membership. Properties of symmetry group theory allow the design of compact probabilistic models of meaning over which inference is efficient and that can be assimilated into a TPKB. We have begun implementing a symmetry-based semantic parser that will learn a semantic symmetry group in an unsupervised way from text.

EXCHANGEABLE VARIABLE MODELS

A sequence of random variables is exchangeable if its joint distribution is invariant under variable permutations. We introduce exchangeable variable models (EVMs) as a novel class of probabilistic models whose basic building blocks are partially exchangeable sequences, a generalization of exchangeable sequences. We present conditions that imply tractable probabilistic inference and discuss parameter and structure learning. We prove that a family of tractable EVMs is optimal under zero-one loss for a large class of functions, including parity and threshold functions, and strictly subsumes existing tractable independence-based model families. Extensive experiments show that EVMs outperform state of the art classifiers such as SVMs and probabilistic models which are solely based on independence assumptions.
This year, we proposed exchangeable variable models (EVMs), a novel family of probabilistic models for classification and probability estimation. While most probabilistic models are built on the notion of conditional independence and its graphical representation, EVMs have finite partially exchangeable sequences as basic components. We show that EVMs can represent complex positive and negative correlations between large sets of variables with few parameters and without sacrificing tractable inference. The parameters of EVMs are estimated under the maximum-likelihood principle and we assume the examples to be independent and identically distributed. We develop methods for efficient probabilistic inference, maximum-likelihood estimation, and structure learning.
We also introduced the mixtures of EVMs (MEVMs) family of models which is strictly more expressive than the naive Bayes family of models but as efficient to learn. MEVMs represent classifiers that are optimal under zero-one loss for a large class of Boolean functions including parity and threshold functions. Extensive experiments show that exchangeable variable models, when combined with the notion of conditional independence, are effective both for classification and probability estimation. The MEVM classifier significantly outperforms state of the art classifiers on numerous high-dimensional and sparse data sets. MEVMs also outperform several tractable graphical model classes on typical probability estimation problems while being orders of magnitudes more efficient.

RELATIONAL SUM-PRODUCT NETWORKS

Relational Sum-Product Networks: Sum-product networks (SPNs; Poon and Domingos 2011) are a recently-proposed deep architecture that guarantees tractable inference, even on certain high-treewidth models. SPNs are a propositional architecture, treating the instances as independent and identically distributed. Last year, we developed Relational Sum-Product Networks (RSPNs), a new tractable first-order probabilistic architecture. RSPNs generalize SPNs by modeling a set of instances jointly, allowing them to influence each other's probability distributions, as well as modeling the probabilities of relations between objects. This year, we developed LearnRSPN, the first algorithm for learning high-treewidth tractable statistical relational models. LearnRSPN is a recursive top-down structure learning algorithm for RSPNs, based on the LearnSPN (Gens and Domingos 2013) algorithm for propositional SPN learning. In our empirical evaluation, the RSPN learning algorithm outperforms Markov Logic Networks (Richardson and Domingos 2006) in both running time and predictive accuracy.

TRACTABLE PROBABILISTIC PROGRAMS

We also developed Tractable Probabilistic Programs (TPP), a statistical relational representation that captures a probability distribution over programs drawn from a context-free grammar. Although TPP was motivated by the problem of automated software debugging, we are investigating its applicability to other problem domains that make use of probabilistic grammars, such as natural language processing and image understanding.

LEARNING CUTSET NETWORKS

Learning tractable probabilistic models from data has been the subject of much recent research. These models offer a clear advantage over Bayesian networks and Markov networks: exact inference over them can be performed in polynomial time,

obviating the need for unreliable, inaccurate approximate inference, not only at learning time but also at query time. Interestingly, experimental results in numerous recent studies have shown that the performance of approaches that learn tractable models from data is similar or better than approaches that learn Bayesian and Markov networks from data. These results suggest that controlling exact inference complexity is the key to superior end-to-end performance.
To take advantage of these promising results, in a recent work that will appear in ECML 2014, we introduced a new tractable probabilistic model called cutset networks for representing large, multi-dimensional discrete distributions. Cutset networks are rooted OR search trees, in which each OR node represents conditioning of a variable in the model, with tree Bayesian networks (Chow-Liu trees) at the leaves. From an inference point of view, cutset networks model the mechanics of Pearl's cutset conditioning algorithm, a popular exact inference method for probabilistic graphical models. We developed efficient algorithms, which leverage and adopt vast amount of research on decision tree induction for learning cutset networks from data. We also developed an expectation-maximization (EM) algorithm for learning mixtures of cutset networks. Our experiments on a wide variety of benchmark datasets clearly demonstrated that compared to approaches for learning other tractable models such as thin-junction trees, latent tree models, arithmetic circuits and sum-product networks, our approach is significantly more scalable, and provides similar or better accuracy on all datasets. In fact, it was better than the competing seven state-of-the-art algorithm on 55\% of the datasets. We are quite excited about these results because the algorithm is quite fast (has provably small computational complexity) and achieves state-of-the-art performance. This makes it an ideal candidate for learning in “Big data” domains.

COMBINING MARKOV LOGIC AND SUPPORT VECTOR MACHINES FOR EVENT EXTRACTION

Event extraction is the task of extracting and labeling all instances in a text document that correspond to a predefined event type. This task is quite challenging because of a multitude of reasons: events are often nested, recursive and have several arguments; there is no clear distinction between arguments and events; etc. For instance, consider the BioNLP Genia shared task Nedellec et al. (2013). In this task, participants are asked to extract instances of a predefined set of Biomedical events from text. An event can have an arbitrary number of arguments that correspond to predefined argument types, and is identified by a keyword called the trigger. The task is complicated by the fact that an event may serve as an argument of another event (nested events). We made the following contributions.
First, we proposed a novel model for Biomedical event extraction based on MLNs that leverages the power of support vector machines (SVMs) Joachims (1999); Vapnik (1995) to handle high-dimensional features. Specifically, we (1) learned SVM models using rich linguistic features for trigger and argument detection and type labeling; (2) designed an MLN composed of soft formulas (each of which encodes a soft constraint whose associated weight indicates how important it is to satisfy the constraint) and hard formulas (constraints that always need to be satisfied, thus having a weight of 1) to capture the relational dependencies between triggers and arguments; and (3) encoded the SVM output as prior knowledge in the MLN in the form of soft formulas, whose weights are computed using the confidence values generated by the SVMs. This formulation naturally allows SVMs and MLNs to complement each other's strengths and weaknesses: learning in a large and sparse feature space is much easier with SVMs than with MLNs, whereas modeling relational dependencies is much easier with MLNs than with SVMs.
Our second contribution concerns making inference with this MLN feasible. Recall that inference involves detecting and assigning the type label to all the triggers and arguments. We showed that existing Maximum-a-posteriori (MAP) inference methods, even the most advanced approximate ones (e.g., Selman et al. (1996), Sontag and Globerson (2011), Marinescu and Dechter (2009), etc.), are infeasible on our proposed MLN because of their high memory cost. To combat this, we identified decompositions of the MLN into disconnected components and solved each independently, thereby drastically reducing the memory requirements.
We evaluated our approach on the BioNLP 2009, 2011 and 2013 Genia shared task datasets.
On the BioNLP'13 dataset, our model significantly outperforms state-of-the-art pipeline approaches and achieves the best F1 score to date. On the BioNLP'11 and BioNLP'09 datasets, our scores are slightly better and slightly worse respectively than the best reported results. However, they are significantly better than state-of-the-art MLN-based systems. A paper on this work will appear at the 2014 Empirical Methods in Natural Language Processing
(EMNLP) conference.

LEARNING BAYESIAN LOGIC PROGRAMS FOR TEXTUAL INFERENCE

The aim of this on-going part of the project is to automatically learn Bayesian Logic Programs (BLPs) from information extracted from natural-language text and use the resulting probabilistic model to make accurate "abductive" inferences from facts extracted from future documents.
We participated in the NIST KBP (Knowledge-Based Population) slot-filling task by using a BLP developed for the KBP ontology to make inferences from text extractions with the goal of increasing recall. We used the publicly distributed version of the CUNY BLENDER system as the base-level KBP extractor. During testing, we used a learned BLP to infer additional facts from the facts extracted by BLENDER, and submitted two sets of results for the competition, one with inferred relations added as well as a baseline set of results without BLP inferences. In order to assemble a large training set for learning a BLP appropriate for KBP, we mapped 26 of the 41 predicates in the KBP ontology to relations in the open-linked database, DBPedia. We then used our previously developed on-line BLP rule learner to learn a BLP from 912,375 mapped facts from DBpedia. For example, one

learned rule was: "If person B is a key employee of organization A, then B is probably a shareholder in A." Unfortunately, partly because the KBP evaluation is focused on evaluating the extraction of explicitly-stated facts rather than probable inferences, the BLP inferences failed to improve recall and actually resulted in an overall decrease in F-measure (form 0.123 to 0.108). In our officially submitted results, we preferred inferred slot fillers to explicitly extracted ones in order to emphasize the role of inference. Subsequent to the official evaluation, we conducted an additional experiment in which we preferred inferred fillers to extracted ones only if their estimated confidence was higher. This version generated 7 additional fillers that were judged correct, resulting in an increase in recall (from .079 to .085) with only a minor decrease in F-measure (from 0.123 to 0.121). This result provides evidence for the value of BLP textual inference despite the limitations of the KBP evaluation with respect to evaluating this capability.
Our recent work has focused on scaling our BLP learning and inference methods to large-scale linked open data, specifically DBPedia. The goal is to learn a BLP from such large-scale data, map the ontology to that for a particular text-extraction task (e.g. KBP), and then use the BLP to make inferences from initial information extracted from text. In order to scale BLP learning to large multi-relational databases such as DBPedia, we have adapted the rule-learning algorithm of Ni Lao at al. (EMNLP, 2011, 2012) to learn the initial relational rules. We then use a simple, approximate maximum-likelihood parameter-learning method we have developed for conditional probability tables (CPTs) that use noisy-or and noisy-and to learn a BLP based on these rules. In order to scale inference to the large, complex BLPs learned from such data, we exploit a semantic-web-based implementation of DataLog, called JENA, to support logical inference, and Gogate's SampleSearch method for efficient and effective probabilistic inference for graphical models with both deterministic and probabilistic constraints.
We have also finalized our plans for evaluating the learned BLPs using the data available in DBPedia, and are currently in the process of conducting a full-scale experimental evaluation. We are using cross-validation on DBPedia data to directly evaluate to the accuracy of BLP-derived inferences. For each fact in a subset of DBPedia, we delete the fact from the database and attempt to infer a value for the corresponding slot using the learned BLP. For example, if we delete the fact that Natasha Obama is a child of Barack Obama, we may be able to infer it from the fact that Natasha is Malia Obama's sister and that Malia is a child of Barack Obama. By using the probability computed using the BLP model to rank the inferred fillers of a slot, we are generating an average precision-recall curve and computing the Mean Average Precision (MAP) to evaluate the accuracy of inference. By comparing the results of BLP inference to that of a purely logical approach (which is unable to meaningfully rank inferred fillers), we plan to measure the advantage of the BLP approach. Preliminary experiments using this methodology have demonstrated promising results, and we are in the process of completing comprehensive experiments using cross-validation on DBPedia. This work will continue as part of the DARPA DEFT project.

DISTRIBUTIONAL MARKOV LOGIC SEMANTICS

The goal of this aspect of the project is to develop an approach to representing the meaning of natural-language sentences as rich, formal expressions in probabilistic logic. An initial logical form is obtained by parsing a sentence using Combinatory Categorial Grammar (CCG). Next, uncertain, distributional information is added as weighted inference rules. The result is a "deep" representation of semantics that captures both logical structure as well as probabilistic, distributional meaning of words and phrases. This representation then supports rich "abductive" probabilistic inference from natural-language text using both Markov logic and Probabilistic Soft Logic (PSL). In particular, we have evaluated the approach on two standard textual inference problems, Recognizing Textual Entailment (RTE) and Semantic Textual Similarity (STS).
Recently, we have worked on improving the efficiency and accuracy of MLN inference for natural-language semantics. We have also explored the use of Probabilistic Soft Logic (PSL) for the STS task.
In March 2014, we participated in Task 1 of SemEval (Semantic Evaluation Workshop): "Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and entailment". The task involved both RTE and STS subtasks on the SICK dataset (Sentences Involving Compositional Knowledge, Marelli et al., to appear). We obtained an accuracy of 73% on RTE, and a Pearson correlation of 0.71 on STS.
Markov Logic Networks can handle all of first-order logic, and have a principled basis in probabilistic logic; however, the networks can grow very large, leading to intractable inference. We have integrated a new inference algorithm based on SampleSearch into Alchemy (the MLN inference system that we are using) to improve run time. We also introduced a modified closed-world assumption that significantly reduces the size of the ground network, thereby making inference feasible. This step has the added benefit of removing extraneous literals from the system, thereby making inference more accurate. Evaluation on the training portion of the SICK RTE data yielded an accuracy of 71.8% for the modified system (original system: 56.9%) with an average runtime of 7s per datapoint (original system: 2min 27s).
We have also explored Probabilistic Soft Logic (Boecheler, Mihalkova and Getoor 2010) as an alternative framework for probabilistic inference for the STS task. We changed the interpretation function for conjunction in PSL to a weighted average to make it more appropriate for STS. In addition, we implemented a new heuristic variant of the lazy grounding implemented in PSL designed to work with the changed implementation of conjunction in a way that avoids the construction of irrelevant groundings. We obtained Pearson correlations of 0.79 on the MSR video corpus, 0.53 on the MRS paraphrase corpus, and 0.71 on the training portion of the SICK STS dataset. In addition, inference was an order of magnitude faster with PSL than with MLNs.
The SICK RTE data allows for three judgments on whether the Text (T) entails the Query (Q): either Entailment, Contradiction, or Neutral. In order to model this three-way distinction, we computed two probabilities, P(Q|T) and P(Q|not(T)), and used a supervised classifier to choose a judgment based on these two probabilities. This setup has the added benefit of addressing the

fundamental problem of MLNs that the computed probability of a sentence depends on both the domain size and the size of the sentence.

Our model combines deep semantics through logical form with weighted inference rules derived from distributional models and can be viewed as an approach to Semantic Parsing that, instead of using a fixed, manually created ontology to interpret predicates, interprets predicate symbols using distributional rules that are automatically created "on the fly."

ROBUST STRUCTURED PREDICTION THROUGH REGULARIZATION

In previous work, we developed max-margin learning methods for collective classification that are robust to adversarial manipulation of object features (Torkamani & Lowd, 2013). However, these methods were restricted to associative Markov networks and could not handle more complex scenarios, such as adversaries that manipulate link structure. We developed a new strategy for learning robust Markov networks or structural SVMs by showing that robustness to perturbations of the features is equivalent to regularization. Specifically, when perturbations are constrained by a norm, the equivalent regularizer is given by the dual norm. When perturbations are constrained by a polyhedron, the equivalent regularizer is a linear function in a transformed space. In experiments, we demonstrate that this regularization strategy leads to improved generalization on a collective classification problem with a lot of concept drift.

DISTRIBUTED GRAPHLAB

While high-level data parallel frameworks, like MapReduce, simplify the design and implementation of large-scale data processing systems, they do not naturally or efficiently support many important data mining and machine learning algorithms and can lead to inefficient learning systems. To help fill this critical void, we introduced the GraphLab abstraction which naturally expresses asynchronous, dynamic, graph-parallel computation while ensuring data consistency and achieving a high degree of parallel performance in the shared-memory setting. We extended the GraphLab framework to the substantially more challenging distributed setting while preserving strong data consistency guarantees. We developed graph based extensions to pipelined locking and data versioning to reduce network congestion and mitigate the effect of network latency. We also introduced fault tolerance to the GraphLab abstraction using the classical Chandy-Lamport snapshot algorithm and demonstrate how it can be easily implemented by exploiting the GraphLab abstraction itself. Finally, we evaluated our distributed implementation of the GraphLab abstraction on a large Amazon EC2 deployment and show 1-2 orders of magnitude performance gains over Hadoop-based implementations.

PARALLEL LEARNING FOR GRAPHICAL MODELS

Two strong areas of focus have been in graphical models and parallel learning. To address these problems in a more accurate fashion, we’ve developed a gradient boosting algorithm for tree-shaped conditional random fields (CRF). Conditional random fields are an important class of models for accurate structured prediction, but effective design of the feature functions is a major challenge when applying CRF models to real world data. Gradient boosting, which can induce and select functions, is a natural candidate solution for the problem. However, it is non-trivial to derive gradient boosting algorithms for CRFs, due to the dense Hessian matrices introduced by variable dependencies. We address this challenge by deriving a Markov Chain mixing rate bound to quantify the dependencies, and introduce a gradient boosting algorithm that iteratively optimizes an adaptive upper bound of the objective function. The resulting algorithm induces and selects features for CRFs via functional space optimization, with provable convergence guarantees. Experimental results on three real world datasets demonstrate that the mixing rate based upper bound is effective for training CRFs with non-linear potentials.

GRAPHLAB: CODE RELEASE AND TECHNOLOGY TRANSFER

One of the major goals of this project is the development of open-source software and of a community around it. We have held two GraphLab workshops in the last couple of years. The first one in 2012 had 318 people in attendance. The second one in 2013 had 570 people. All our code is available at http://graphlab.org. The GraphLab open-source project, started by PI Guestrin, has received very significant attention in industry and academia. As discussed above, the software has received tens of thousands of downloads, and held two very popular workshops. This project has had very significant impact in industry and academia. To continue to support the users of GraphLab, and to continue to expand its reach, we have recently spun off a company, where GraphLab can become its own entity beyond the university. This company has recently announced its first round of funding, receiving $6.75M.

PROGRESS ON ACTIVITY AND PLAN RECOGNITION

Activity recognition is central to many important problems including surveillance (recognizing the activities of an opponent), anomaly detection (recognizing and ignoring normal behaviors), and human-computer interaction (recognizing the goals and activities of the user). Activity recognition algorithms seek to infer the goals and plans of one or more agents from noisy and fragmentary observations of their behavior. This is a classic problem of abductive inference.

ACTIVITY RECOGNITION IN THE KITCHEN

In our first application, we implemented and evaluated a Markov-logic based plan recognition system for kitchen activities, where observations came from video and natural language narration (Song et al 2013). We presented a general framework for complex event recognition that is well-suited for integrating information that varies widely in detail and granularity. Consider the scenario of an agent in an instrumented space performing a complex task while describing what he is doing in a natural manner. The system takes in a variety of information, including objects and gestures recognized by RGB-D and descriptions of events extracted from recognized and parsed speech. The system outputs a complete reconstruction of the agent's plan, explaining actions in terms of more complex activities and filling in unobserved but necessary events. We show how to use Markov Logic (a probabilistic extension of first-order logic) to create a model in which observations can be partial, noisy, and refer to future or temporally ambiguous events; complex events are composed from simpler events in a manner that exposes their structure for inference and learning; and uncertainty is handled in a sound probabilistic manner.
We evaluated our framework on a multi-modal corpus collected from people conducting tasks in an instrumented kitchen, including making tea, making cocoa and making oatmeal. Participants were asked to conduct the activity and at the same time verbally describe the action being conducted. The experiments demonstrated that (i) employing a complex event library improves visual event detection, and (ii) using both an event library and data from free-form spoken language can compensate for sparse visual input.

ACTIVITY RECOGNITION IN SOCIAL MEDIA

We extended our work on activity and state recognition from social media data (Sadelik et al 2013, Brennan et al 2013). Computational approaches to health monitoring and epidemiology continue to evolve rapidly. We presented an end-to-end system, nEmesis, that automatically identifies restaurants posing public health risks. Leveraging a language model of Twitter users" online communication, nEmesis finds individuals who are likely suffering from a foodborne illness. People"s visits to restaurants are modeled by matching GPS data embedded in the messages with restaurant addresses. As a result, we can assign each venue a "health score" based on the proportion of customers that fell ill shortly after visiting it. Statistical analysis reveals that our inferred health score correlates (r = 0:30) with the official inspection data from the Department of Health and Mental Hygiene (DOHMH). We investigated the joint associations of multiple factors mined from online data with the DOHMH violation scores and find that over 23% of variance can be explained by our factors. We demonstrated that readily accessible online data can be used to detect cases of foodborne illness in a timely manner. This approach offers an inexpensive way to enhance current methods to monitor food safety (e.g., adaptive inspections) and identify potentially problematic venues in near-real time.

MAKING MODEL MARKOV LOGIC MORE EFFICIENT

We developed complexity results on a "less intractable" subset of multi-agent Markov Logic (Papai & Kautz 2013). Modal Markov Logic for a single agent has previously been proposed as an extension to propositional Markov logic. While the framework allowed reasoning under the principle of maximum entropy for various modal logics, it is not feasible to apply its counting based inference to reason about the beliefs and knowledge of multiple agents due to magnitude of the numbers involved. We propose a modal extension of propositional Markov logic that avoids this problem by coarsening the state space. The problem stems from the fact that in the single-agent setting, the state space is only doubly exponential in the number of propositions in the domain, but the state space can potentially become infinite in the multi-agent setting. In addition, the proposed framework adds only the overhead of deciding satisfiability for the chosen modal logic on the top of the complexity of exact inference in propositional Markov logic. The proposed framework allows one to find a distribution that matches probabilities of formulas obtained from training data (or provided by an expert). Finally, we showed how one can compute lower and upper bounds on probabilities of arbitrary formulas.

PLAN RECOGNITION WITH MONTE CARLO TREE SEARCH

We continued investigating the application of Monte Carlo tree search (MCTS algorithms to planning and plan recognition. This year, we developed the mathematical foundations for state abstraction in MCTS. In previous work, we pioneered formal methods for temporal and state abstraction in hierarchical reinforcement learning (HRL). The requirements for correct state abstraction in HRL are very stringent and, hence, rarely satisfied in practice. In contrast, we showed that state abstraction in MCTS is much easier to achieve. We proved accuracy bounds for a certain form of state abstraction, state aggregation in ExpectiMax trees, and we showed that these state abstractions preserve optimality in search trees. This in turn permitted us to prove correctness of state aggregation abstractions for two MCTS methods: Sparse Sampling and UCT. These results are very general and show excellent performance improvements in several benchmark problems. Future work will focus on automatically learning these state abstractions.

TECHNOLOGY TRANSFER

The software and methods developed as part of the project have found numerous applications both in industry and academia. In the following, we list some of these technology transfers.

DARPA PPAML (Probabilistic Programming for Advanced Machine Learning):
The project was launched in Fall 2013, and developed in part based on research funded under this MURI in the Tenenbaum group. An additional MURI member (Dietterich) is playing a crucial role on the PPAML evaluation team.
Several teams of the project are using algorithms in Alchemy 2.0 for building their probabilistic programming systems as well as for competing in the DARPA evaluations.

Vibhav Gogate along with Avi Pfeffer from Charles Rivers Analytics received a Phase 1 AFOSR SBIR grant on “Representation and Inference for Developing Deep Language Engines (RIDDLE).” The project used lifted algorithm in Alchemy 2.0 for solving complex NLP tasks such as Event Extraction and temporal relation classification.

Daniel Lowd received a Google faculty research award funding work on tractable probabilistic models, research that was initiated as part of the MURI project.
Tractable Markov logic and Tractable Probabilistic Knowledge Bases are applied and extended by Domingos’ group within the DARPA DEFT (Deep Exploration of Text) project and the ONR BRC project on Structured Learning for Scene Understanding.

Tenenbaum helped to organize and keynote a workshop on the interface between computational cognitive modeling and the Intelligence Community, sponsored by IARPA and BBN/Raytheon. This workshop was attended by approximately 80 members of the IC and associated research agencies, at Raytheon offices near Fort Meade. MURI funded research was presented and generated significant interest and follow-up from multiple attendees.

Core research from this grant (the GraphLab system) was spun off as a company. This start up received $6.75M in VC funding, and is currently employing 26 people.

COMPANIES AND INDIVIDUALS WHO HAVE WORKED WITH ALCHEMY

Google Inc
Kevin Murphy (murphyk@cs.ubc.ca)
Brian Milch (milch@google.com)

Microsoft Research
Past Contributors to Alchemy already at Microsoft research (Matt Richardson, Hoifung Poon)
Ben Livshits (livshits@microsoft.com)
many others

IBM Research
Ashish Sabharwal (now at AI2; Paul Allen Institute in Seattle)

LogicBlox
Benny Kimelfeld (bennyk@gmail.com)
Molham Aref (molham.aref@logicblox.com)

Charles Rivers Analytics
Avi Pfeffer (apfeffer@cra.com)

Facebook
Kedar Bellare (kedar.bellare@gmail.com)

Baylor Hospital Network
Dr. Jack Stecher (jackstecher@sbcglobal.net)

Amazon
Srikanth Doss (srikanth.doss@gmail.com)

Netflix
Qi Zhong (no longer at Netflix)

Yahoo
Andrew Gelfand (agelfand@ics.uci.edu)
Vibhor Rastogi (no longer works at Yahoo; now at Facebook)

BAE systems
Gregory Sullivan (gregory.sullivan@baesystems.com)

Raytheon
Kenric P Nelson (Kenric_P_Nelson@raytheon.com)

SRI International
Rodrigo de Salvo Braz (braz@ai.sri.com)

Samsung
Jae Hyun Son (jhson@sta.samsung.com)

Match.com
Vaclav Patricek (patricek@match.com)

BBN technology
Hala Mostafa (hmostafa@bbn.com)

CycCorp
Michael Witbrock (witbrock@cyc.com)

Gamelan systems
Ben Vigoda (ben.vigoda@gamelanlabs.com)

Analog Devices
Theo Weber (theo.weber@analog.com)

Intel
Denver Dash (now a research Professor at CMU)
Many others affiliated with center at UW

Nuance Communications
Hung Bui (bui.h.hung@gmail.com)

Ebay
Ravi Jammalamadaka (ravij@ebay.com)

Hitachi
Robert Mateescu (mateescu@hitachi.com)

Other companies which have used Alchemy but we do not have contacts for At&T, Nokia, Twitter, Xerox Corp

REFERENCES

Bahar, R., Frohm, E., Gaona, C., Hachtel, G., Macii, E., Pardo, A., and Somenzi, F. (1993). Algebraic decision diagrams and their applications. In Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, pages 188-191.

Bonawitz, E. and Ullman, T. and Gopnik, A. and Tenenbaum, J.B. (2012). Sticking to the evidence? A computational and behavioral case study of micro-theory change in the domain of magnetism. Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics.

Cheng, J. and Druzdzel, M. J. (2000). AIS-BN: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. Journal of Artificial Intelligence Research, 13:155-188.

Chol Song, Y. and Kautz, H. (2012). A Testbed for Learning by Demonstration from Natural Language and RGB-Depth Video. Proceedings of the 26th Conference on Artificial Intelligence.

Darwiche, A. (2003). A Differential Approach to Inference in Bayesian Networks. Journal of the ACM, 50:280-305.

Diaconis, Persi and Freedman, David. De Finetti’s generalizations of exchangeability. In Studies in Inductive Logic and Probability, volume II. 1980a.

Diaconis, Persi and Freedman, David. Finite exchangeable sequences. The Annals of Probability, 8(4):745–764, 1980b.

Dechter, E. and Malmaud, J. and Adams, R.P. and Tenenbaum, J.B. (2013). Bootstrap learning via modular concept discover. Proceedings of the International Joint Conference on Artificial Intelligence.

Domingos, P. and Lowd, D. (2009). Markov Logic: An Interface Layer for Artificial Intelligence. Morgan & Claypool, San Rafael, CA.

Domingos, P. and Webb, A. (2012) A Tractable First-Order Probabilistic Logic. Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.

Domingos, P. and Webb, A. (2012). A tractable First-order probabilistic logic. Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.

Duvenaud, D. and Lloyd, J.R. and Grosse, R. and Tenenbaum, J.B. and Ghahramani, Z. (2013). Structure Discovery in Nonparametric Regression through Compositional Kernel Search. Proceedings of the International Conference on Machine Learning.

Elidan, G. and Globerson, A. (2010). The 2010 UAI approximate inference challenge. Available online at: http://www.cs.huji.ac.il/project/UAI10/index.php.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721-741.

Gens, R. and Domingos, P. (2012) Discriminative Learning of Sum-Product Networks. Advances in Neural Information Processing Systems 25.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). Markov Chain Monte Carlo in Practice. Chapman and Hall, London, UK.

Gogate, V. and Dechter, R. (2007). SampleSearch: A scheme that Searches for Consistent Samples. Proceedings of the 11th Conference on Artificial Intelligence and Statistics (AISTATS), pages 147-154.

Gogate, V. and Dechter, R. (2011). SampleSearch: Importance sampling in presence of determinism. Artificial Intelligence, 175 (2):694-729.

Gogate, V. and Domingos, P. (2011). Probabilistic Theorem Proving. Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence.

Gonzalez, J. and Low, Y. and Gu, H. and Bickson, D. and Guestrin, C. (2012). PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs. Proceedings of Operating Systems Design and Implementation.

Grosse, R.B. and Salakhutdinov, R. and Freeman, W.T. and Tenenbaum, J.B. (2012). Exploiting compositionality to explore a large space of model structures. Proceedings of the Conference on Uncertainty in Artificial Intelligence.

Jara-Ettinger, J. and Baker, C.L. and Tenenbaum, J.B. (2012). Learning What is Where from Social Observations. Proceedings of the 34th Annual Conference of the Cognitive Science Society.

Jensen, C. S., Kjaerulff, U., and Kong, A. (1993). Blocking Gibbs Sampling in Very Large Probabilistic Expert Systems. International Journal of Human Computer Studies. Special Issue on Real-World Applications of Uncertain Reasoning, 42:647-666.

Kyrola, A., and Blelloch, G. and Guestrin, C. (2012). GraphChi: Large-Scale Graph computation on Just a PC. Proceedings of the 10th USENIX Symposium on Operating Systems Design and Implementation.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. Journal of the Royal Statistical Society. Series B (Methodological), 50(2):157-224.

Liu, J. S., Wong, W. H., and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. Biometrika, 81:27-40.

Low, Y. and Gonzalez, J. and Kyrola, A. and Bickson, D. and Guestrin, C. and Hellerstein, J.M. (2012). Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud. Proceedings of Very Large Data Bases.

Swift, M. and Ferguson, G. and Galescu, L., and Chu, Y. and Harman, C. and Jung, H. and Perera I. and Chol Song, Y. and Allen J. and Kautz, H. (2012). A multimodal corpus for integrated language and action. Proceedings of the International Workshop on Multimodal Corpora for Machine Learning.

Poon, H and Domingos, P. (2011) Sum-Product Networks: A New Deep Architecture. Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence.

Sarkhel, S. and Gogate, V. (2013). Lifting walksat-based local search algorithms for map inference. Proceedings of the 3rd International Workshop on Statistical Relational Artificial Intelligence.

Smith, D. and Gogate, V. (2013). The inclusion-exclusion rule and its application to the junction tree algorithm. Proceedings of the International Joint Conference on Artificial Intelligence.

Venugopal, D. and Gogate, V. (2013a). Dynamic blocking and collapsing for gibbs sampling. Proceedings of the Conference on Uncertainty in Artificial Intelligence.

Venugopal, D. and Gogate, V. (2013b). Giss: Combining SampleSearch and importance sampling for inference in mixed probabilistic and deterministic graphical models. Proceedings of the AAAI Conference on Artificial Intelligence.

Webb, W.A. and Domingos, P. (2013). Tractable Probabilistic Knowledge Bases with Existence Uncertainty. Proceedings of the 3rd International Workshop on Statistical Relational Artificial Intelligence.

Technology Transfer

DARPA PPAML (Probabilistic Programming for Advanced Machine Learning):
The project was launched in Fall 2013, and developed in part based on research funded under this MURI in the Tenenbaum group. An additional MURI member (Dietterich) is playing a crucial role on the PPAML evaluation team.
Several teams of the project are using algorithms in Alchemy 2.0 for building their probabilistic programming systems as well as for competing in the DARPA evaluations.

Vibhav Gogate along with Avi Pfeffer from Charles Rivers Analytics received a Phase 1 AFOSR SBIR grant on “Representation and Inference for Developing Deep Language Engines (RIDDLE).” The project used lifted algorithm in Alchemy 2.0 for solving complex NLP tasks such as Event Extraction and temporal relation classification.

Daniel Lowd received a Google faculty research award funding work on tractable probabilistic models, research that was initiated as part of the MURI project.
Tractable Markov logic and Tractable Probabilistic Knowledge Bases are applied and extended by Domingos’ group within the DARPA DEFT (Deep Exploration of Text) project and the ONR BRC project on Structured Learning for Scene Understanding.

Tenenbaum helped to organize and keynote a workshop on the interface between computational cognitive modeling and the Intelligence Community, sponsored by IARPA and BBN/Raytheon. This workshop was attended by approximately 80 members of the IC and associated research agencies, at Raytheon offices near Fort Meade. MURI funded research was presented and generated significant interest and follow-up from multiple attendees.

Core research from this grant (the GraphLab system) was spun off as a company. This start up received $6.75M in VC funding, and is currently employing 26 people.

Keywords: Markov Logic, Abductive Inference, Tractable Probabilistic Knowledge Bases, Lifted Probabilistic Inference, Symmetry-based Learning and Inference, Non-convex Optimization, Combining Markov Logic and Support Vector Machines, Textual Inference, Adversarial Collective Classification, Monte Carlo Tree Search, Activity Recognition, Event Extraction, Plan Recognition, Large-Scale Parallel Learning.

ABSTRACT

The project's main focus was on tractable inference and learning of probabilistic representations, which are essential for large-scale abductive inference applications. We also developed novel inference techniques based on lifting, sampling, and more efficient processing of evidence. We continued to extend Alchemy 2.0, an open-source toolkit for Markov logic, and Alchemy Lite, an implementation of Tractable Markov Logic (TML). We developed parameter and structure learning algorithms for sum-product networks and, building on TML, we substantially improved two tractable probabilistic-logical formalisms: relational sum-product networks and tractable probabilistic knowledge bases. Based on sum-product networks, we worked towards formalisms for tractable probabilistic programming. We worked on symmetry-based inference and learning and developed novel model classes that exploit invariances of the data with respect to group operations. A novel model for Biomedical event extraction based on MLNs that leverages the power of support vector machines (SVMs) to handle high-dimensional features was proposed and applied to the problem of event extraction. We developed structured prediction models by introducing novel forms of regularization. We continued to apply Markov logic networks to the problem of textual inference and conducted extensive experiments on benchmark datasets. We further improved GraphLab, our large-scale parallel machine learning framework. We investigated novel approaches to activity and plan recognition, and showed that Markov logic is capable of fusing visual and language evidence of the activities under consideration.

INTRODUCTION AND PROGRESS OVERVIEW

The goal of this MURI project is to develop a unified approach to abductive inference, which combines the capabilities of logic and probability. Abduction is inference to the best explanation. Consider the situation of a military commander. He needs to make sense of a bewildering array of information: sensors carried by troops on patrol, sensors placed along roads, video from surveillance cameras, video and other feeds from unmanned vehicles (aerial and ground), remote sensing streams, intercepted communications, mission reports, intelligence reports, news stories, etc. Critical decisions depend on the correct interpretation of this data: Where to send troops? Should a convoy be rerouted? Is an attack imminent? Is the behavior of a group of people suspicious? Bridging the gulf between the mass of low-level sensor information and the necessary high-level understanding of a situation requires abductive inference. Humans are experts at this, but they can only handle so much information at a time (and also have well-documented biases and blind spots). Making the most of the available information requires automated inference that continuously supports and complements the commander. However, the scale, uncertainty and complexity of the task puts it beyond the reach of current AI systems.

An effective abductive inference system has to interpret evidence, construct explanations, discard irrelevant data (but revisit it if becomes relevant), integrate information from many different sources, formulate and test hypotheses, suggest alternative courses of action, and help direct the collection of further data. It must be able to reason at every scale, from interpreting the moment-by-moment actions of the enemy and the population it is embedded into detecting plans whose steps may be far apart in time (e.g., scoping a target location, procuring materials for a bomb, assembling it, deploying it, etc.), to understanding the motivations and thought processes of the enemy.

The goal of the research proposed here is to make such a system a reality. We are developing a well-founded approach to abduction based on the solid foundations of first-order logic and probability. We are using Markov logic (Domingos and Lowd, 2009), which unifies first-order logic and probability, as the common representation language.

The research program we proposed has four major components: foundations, inference, learning, and activity and plan recognition. We completed the proposed research on foundations in the first three years. In the following, we will describe our accomplishments and progress made in the past reporting period on the latter three components, starting with a brief overview.

INFERENCE

This component of the project seeks to build scalable, next-generation inference systems for Markov logic but also particular variations of Markov logic that render inference and/or learning more tractable.

We have made progress in the following areas:

- Fast Evidence Processing in Markov Logic: Evidence breaks symmetries and lifted inference algorithms, state of the art inference methods for statistical relational models, end up grounding the MLN. To solve these problems, we have developed a general method to achieve scalable inference in MLNs, allowing for arbitrary structures with arbitrary evidence. We reduced the size of the inference problem of Markov Logic Networks by clustering together similar evidence atoms, and replacing all atoms in each cluster by their cluster center. To formulate the clustering problem, we have developed several similarity measures. We performed experiments on several different benchmark MLNs utilizing various clustering and inference algorithms. Our experiments clearly show the generality and scalability of our approach.
- Lifted MAP Inference for Markov Logic: We improved existing lifted MAP inference results further by showing that if non-shared MLNs contain no self joins, namely every atom appears at most once in each of its formulas, then all variables in the corresponding

Markov network need only be bi-valued. Our approach is quite general and can be easily applied to arbitrary MLNs by simply grounding all of its shared terms. The key feature of our approach is that because we reduce lifted inference to propositional inference, we can use any propositional MAP inference algorithm for performing lifted MAP inference, without actually lifting the propositional algorithm.

- Loopy Belief Propagation in presence of determinism: We proposed a new method for improving the performance of loopy belief propagation in the presence of logical constraints and determinism. The key idea in our method is nding a reparameterization of the graphical model such that LBP, when run on the reparameterization, is likely to have better convergence properties than LBP on the original graphical model. We proposed several new schemes for nding such reparameterizations, all of which leverage unique properties of zeros as well as research on LBP convergence done over the last decade. Our experimental evaluation on a variety of PGMs clearly demonstrates the promise of our method -- it often yields accuracy and convergence time improvements of an order of magnitude or more over standard LBP.
- Probabilistic Inference for Hybrid MLNs: We have continued to develop an efficient algorithm for continuous, nonconvex optimization, which we call RDIS. RDIS allows us to efficiently perform MAP inference in continuous domains and to accurately fit models with continuous parameters to data. RDIS easily supports discrete variables as well as continuous, enabling it to play an important role in multi-modal and multi-scale domains. The key to RDIS is to identify and exploit local structure in the objective function by dynamically identifying a subset of variables that, once optimized, decomposes the remaining variables into approximately independent subsets. These can then be separately optimized, and we do so recursively in order to exploit multiple levels of local structure, dynamically finding within each a subset that further decomposes it, etc.
- Tractable Probabilistic Knowledge Bases (TPKB): Tractable Markov Logic was originally designed to be used with Probabilistic Theorem Proving (PTP) as an inference algorithm, but PTP was much more complicated than necessary for the restricted case, which clouded intuition and made formal guarantees difficult. Last year, we worked on improvements of TML that feature existence uncertainty and an object oriented syntax. This year, we further improved TPKBs and learned a large-scale TPKB from a variety of data sources, a longstanding goal of AI research. Existing approaches either ignore the uncertainty inherent to knowledge extracted from text, the web, and other sources, or lack a consistent probabilistic semantics with tractable inference. TPKBs consist of a hierarchy of classes of objects and a hierarchy of classes of object pairs such that attributes and relations are independent conditioned on those classes. These characteristics facilitate both tractable probabilistic reasoning and tractable maximum-likelihood parameter learning. TPKBs feature a rich query language that allows one to express and infer complex relationships between classes, relations, objects, and their attributes. The queries are translated to sequences of operations in a relational database facilitating query execution times in the sub-second range. We demonstrate the power of TPKBs by leveraging large data sets extracted from Wikipedia to learn their structure and parameters. The resulting TPKB models a distribution over millions of

objects and billions of parameters. We apply the TPKB to entity resolution and object linking problems and show that the TPKB can accurately align large knowledge bases and integrate triples from open IE projects.

- Alchemy 2.0 and Alchemy Lite: We continued to work on Alchemy (our open source software package for learning and inference in Markov Logic). The main highlight of the newer version is access to highly scalable lifted inference algorithms, lifted sampling approaches developed in this project over the past two years. We also continued to work on Alchemy Lite, an open-source software package for inference in Tractable Markov Logic (TML), the first tractable first-order probabilistic logic. For a list of companies and people that have used Alchemy, we refer the reader to the end of the report.

LEARNING

The goal of learning algorithms is to acquire knowledge from data autonomously, without human intervention. In the past year, we have made progress on the following fronts.

- Structure Learning for Sum-Product Networks: We developed a new SPN structure learning algorithm, called ID-SPN, for learning SPNs with both indirect and direct variable interactions. ID-SPN performs a top-down clustering of instances and variables, similar to our previous method, but with tractable graphical models at the leaves instead of univariate distributions. These leaf distributions are learned using ACMN, a state-of-the-art method we developed for learning arithmetic circuits (Lowd & Rooshenas, 2013). ID-SPN uses the likelihood of a validation set to determine the proper depth of each branch. In experiments on a standard set of benchmark datasets, ID-SPN is more accurate than the previous state-of-the-art on 20 out of 20 datasets. ID-SPN also obtains better likelihoods than intractable Bayesian networks on 13 out of 20 datasets, suggesting that tractable models can be as accurate as intractable ones.
- Sum-Product Networks for Vision: We investigated the use of SPNs for fast object recognition and three-dimensional pose estimation. We continued to investigate the use of Sum-Product Networks for object recognition and pose estimation. As a step toward this goal, we developed the Deep Symmetry Network (DSN), a neural network that can model invariance to richer transformations.
- Deep Symmetry Networks: We developed deep symmetry networks (symnets), a generalization of convnets that forms feature maps over arbitrary symmetry groups. Symnets use kernel-based interpolation to tractably tie parameters and pool over symmetry spaces of any dimension. They also use a Lucas-Kanade optimization to warp features to feature maps. Experiments on the NORB and MNIST-rot datasets show that symnets over the affine group greatly reduce sample complexity relative to convnets by better capturing the symmetries in the data.
- Symmetry-based Semantic Parsing: We have developed a new notion of semantics based on symmetry group theory. Our new approach allows us to develop a semantic parser that avoids the challenges of having to choose a formal meaning representation or

having to gather large amounts of labeled training data; our parser also represents semantics in a way that is easily integrated into a Tractable Probabilistic Knowledge Base over which abductive inference will be performed.

- Exchangeable Variable Models: Conditional independence is a crucial notion that facilitates efficient inference and parameter learning in probabilistic models. Its logical and algorithmic properties as well as its graphical representations have led to the advent of graphical models as a discipline within artificial intelligence. The notion of finite (partial) exchangeability (Diaconis & Freedman, 1980a), on the other hand, has not yet been explored as a basic building block for tractable probabilistic models. A sequence of random variables is exchangeable if its distribution is invariant under variable permutations. Similar to conditional independence, partial exchangeability, a generalization of exchangeability, can reduce the complexity of parameter learning and is a concept that facilitates high tree-width graphical models with tractable inference.
- Relational Sum-Product Networks: Building on our earlier work on Relational Sum-Product Networks, we developed LearnRSPN, the first algorithm for learning high-treewidth tractable statistical relational models. LearnRSPN is a recursive top-down structure learning algorithm for RSPNs, based on the LearnSPN (Gens and Domingos 2013) algorithm for propositional SPN learning. In our empirical evaluation, the RSPN learning algorithm outperforms Markov Logic Networks (Richardson and Domingos 2006) in both running time and predictive accuracy.
- Tractable Probabilistic Programs: We began to develop Tractable Probabilistic Programs (TPP), a statistical relational representation that captures a probability distribution over programs drawn from a context-free grammar. Although TPP was motivated by the problem of automated software debugging, we are investigating its applicability to other problem domains that make use of probabilistic grammars, such as natural language processing and image understanding.
- We proposed a novel model for Biomedical event extraction based on MLNs that leverages the power of support vector machines (SVMs) to handle high-dimensional features. Specifically, we learned SVM models using rich linguistic features for trigger and argument detection and type labeling; designed an MLN composed of soft formulas (each of which encodes a soft constraint whose associated weight indicates how important it is to satisfy the constraint) and hard formulas (constraints that always need to be satisfied, thus having a weight to capture the relational dependencies between triggers and arguments; and encoded the SVM output as prior knowledge in the MLN in the form of soft formulas, whose weights are computed using the confidence values generated by the SVMs. This formulation naturally allows SVMs and MLNs to complement each other's strengths and weaknesses: learning in a large and sparse feature space is much easier with SVMs than with MLNs, whereas modeling relational dependencies is much easier with MLNs than with SVMs.
- BLPs for Textual Inference: During the final year of the project, the research at the University of Texas at Austin has focused on applying and adapting the statistical-relational AI techniques we developed under the MURI project to the problem of abductive inference for natural-language text understanding, which we are investigating

as part of DARPA's Deep Exploratory and Filtering of Text (DEFT) program. We used the Bayesian Logic Programming (BLP) methods we developed to learn knowledge-bases for making probabilistic inferences from information automatically extracted from text.

- Distributional Markov Logic Semantics: We applied some of the Markov logic methods we developed to construct a general formal semantics for natural language that integrates traditional logical form produced by a broad-coverage parser with probabilistic rules extracted from a vector-space distributional semantics automatically constructed from large corpora. These techniques were evaluated on the Knowledge Base Population (KBP) formal evaluation conducted by NIST and on standardized datasets for Recognizing Textual Entailment (RTE) and Semantic Textual Similarity (STS).
- Robust Structured Prediction through Regularization: In previous work, we developed max-margin learning methods for collective classification that are robust to adversarial manipulation of object features (Torkamani & Lowd, 2013). However, these methods were restricted to associative Markov networks and could not handle more complex scenarios, such as adversaries that manipulate link structure. We developed a new strategy for learning robust Markov networks or structural SVMs by showing that robustness to perturbations of the features is equivalent to regularization. Specifically, when perturbations are constrained by a norm, the equivalent regularizer is given by the dual norm. When perturbations are constrained by a polyhedron, the equivalent regularizer is a linear function in a transformed space. In experiments, we demonstrate that this regularization strategy leads to improved generalization on a collective classification problem with a lot of concept drift.
- Large-Scale Machine Learning: We focused on the long term goal of building and improving our GraphLab large scale parallel machine learning framework (http://graphlab.org). We extended the GraphLab framework to the substantially more challenging distributed setting while preserving strong data consistency guarantees. Two strong areas of focus have been in graphical models and parallel learning. To address these problems in a more accurate fashion, we've developed a gradient boosting algorithm for tree-shaped conditional random fields (CRF). Conditional random fields are an important class of models for accurate structured prediction, but effective design of the feature functions is a major challenge when applying CRF models to real world data. Gradient boosting, which can induce and select functions, is a natural candidate solution for the problem. However, it is non-trivial to derive gradient boosting algorithms for CRFs, due to the dense Hessian matrices introduced by variable dependencies. We address this challenge by deriving a Markov Chain mixing rate bound to quantify the dependencies, and introduce a gradient boosting algorithm that iteratively optimizes an adaptive upper bound of the objective function. The resulting algorithm induces and selects features for CRFs via functional space optimization, with provable convergence guarantees. Experimental results on three real world datasets demonstrate that the mixing rate based upper bound is effective for training CRFs with non-linear potentials.

ACTIVITY AND PLAN RECOGNITION

Activity and plan recognition is a classic problem of abductive inference. The goal is to infer the goals and plans of one or more agents from noisy and fragmentary observations of their behaviour. During the reporting period, we made significant progress on both applied and fundamental work on plan and activity recognition:

- Kitchen activities: We implemented and evaluated a Markov-logic based plan recognition system for kitchen activities, where observations came from video and natural language narration.
- Disease Prediction with Social Media Data: We extended our work on activity and state recognition from social media data. We showed that we could model disease transmission at a global scale using posts from airline travellers, and could identify possible sources of food poisoning by identifying restaurant meal events and sickness events from Twitter posts.
- More Efficient Modal Markov Logic: We developed complexity results on a "less intractable" subset of multi-agent Markov Logic.
- Plan Recognition with Monte Carlo Tree Search: We developed the mathematical foundations for state abstraction in MCTS. We showed that state abstraction in MCTS is much easier to achieve. We proved accuracy bounds for a certain form of state abstraction, state aggregation in ExpectiMax trees, and we showed that these state abstractions preserve optimality in search trees. This in turn permitted us to prove correctness of state aggregation abstractions for two MCTS methods: Sparse Sampling and UCT. These results are very general and show excellent performance improvements in several benchmark problems. Future work will focus on automatically learning these state abstractions.

BACKGROUND: MARKOV LOGIC

Markov logic (Domingos and Lowd, 2009) provides the foundation for our research into abductive inference. It is unique in its simplicity and generality, and in the range, scalability and sophistication of its algorithms, which are publicly available in the open-source Alchemy package. Markov logic attaches weights to formulas in first-order logic. A first-order knowledge base can be seen as a set of hard constraints on the set of possible worlds: if a world violates even one formula, it has zero probability. The basic idea in Markov logic is to soften these constraints: when a world violates one formula in the knowledge base it is less probable, but not impossible. The fewer formulas a world violates, the more probable it is. A formula's associated weight reflects how strong a constraint it is: the higher the weight, the greater the difference in log probability between a world that satisfies the formula and one that does not, other things being equal. We call a set of weighted first-order formulas a Markov logic network (MLN). Semantically, we view the formulas as templates for constructing Markov networks (the undirected counterpart of Bayes nets). Given different sets of constants, an MLN will produce different networks, and these may be of widely varying size, but all will have certain regularities in structure and parameters (e.g., all groundings of the same formula will have the same weight). Markov logic has first-order logic and most discrete statistical models used in AI as special cases. Alchemy currently includes algorithms for inferring the most probable explanation of evidence, computing marginal and conditional probabilities, learning formula weights from data (generatively and discriminatively), and learning and/or revising formulas.

BACKGROUND: TRACTABLE MARKOV LOGIC (TML)

We have to date made much progress on combining first-order logic with probability, starting with Markov Logic and more recently with Probabilistic Theorem Proving (PTP) (Gogate and Domingos, 2011). However, both of these approaches suffer from the fact that even simple models created with them are often intractable, which essentially precludes their widespread use. Sum-Product Networks (SPNs) (Poon and Domingos, 2011) have guaranteed tractability, but at the expense of greatly reduced expressiveness, specifically, they are not first-order and can be difficult to interpret. Tractability and expressiveness seem to be fundamentally opposed and one might think that having useful first-order features would necessarily allow for intractable models, however, we have devised a new language that we believe combines substantial first-order representational richness with guaranteed tractability. This language, which we call Tractable Markov Logic (TML) is a subset of Markov Logic whose structure was informed by both SPNs and PTP. TML can represent objects and relations in a first-order fashion and is structured according to ontology-like class and part hierarchies. These restrictions are strong enough to allow for an exact inference algorithm that is linear in the number of objects times the number of rules in the knowledge base. However, they are also weak enough that TML can compactly represent essentially all widely-used tractable models, including junction trees, probabilistic context-free grammars and SPNs. Additionally, TML permits probabilistic versions of inheritance hierarchies and default reasoning. These results are described in our paper (Domingos and Webb, 2012), implemented in preliminary form as a software package called Alchemy Lite, will be presented soon.

PROGRESS ON INFERENCE

LOOPY BELIEF PROPAGATION IN THE PRESENCE OF LOGICAL DEPENDENCIES

It is well known that loopy Belief propagation (LBP), perhaps the most widely used and the most researched inference algorithm, performs poorly on probabilistic graphical models (PGMs) with determinism and logical dependencies. This is problematic because many probabilistic programs contain large amount of determinism and logical constraints. Therefore, in this work, we proposed a new method for remedying this problem. The key idea in our method is finding a reparameterization of the graphical model such that LBP, when run on the reparameterization, is likely to have better convergence properties than LBP on the original graphical model. We proposed several new schemes for finding such reparameterizations, all of which leverage unique properties of zeros as well as research on LBP convergence done over the last decade. Our experimental evaluation on a variety of PGMs clearly demonstrates the promise of our method -- it often yields accuracy and convergence time improvements of an order of magnitude or more over LBP. This work was published at the 2014 AISTATS conference.

FAST EVIDENCE PROCESSING IN MARKOV LOGIC NETWORKS

Markov Logic Networks (MLNs) unify first order logic and probabilistic graphical models. However, due to the rich representational power of MLNs, inference in these models is extremely challenging. The standard graphical model inference algorithms operate on the ground model and do not scale well as the number of objects gets larger. On the other hand, lifted inference algorithms which perform inference at the first-order level offer the desired scalability. However, the conditions under which the MLN can be correctly processed at the lifted level are often very restrictive. A worse problem is that, evidence breaks symmetries and lifted inference algorithms end up grounding the MLN. To solve these problems, we have developed a general method to achieve scalable inference in MLNs, allowing for arbitrary structures with arbitrary evidence.

Our approach works is quite straight-forward. Given a MLN and a large set of evidence atoms, we reduce the size of the inference problem by clustering together similar evidence atoms, and replacing all atoms in each cluster by their cluster center. To formulate the clustering problem, we have developed several similarity measures. We performed experiments on several different benchmark MLNs utilizing various clustering and inference algorithms. Our experiments clearly show the generality and scalability of our approach. This work will appear in the 2014 ECML conference.

LIFTED MAP INFERENCE

We developed a new approach for approximate MAP inference in Markov Logic Networks (MLNs) (this work was published at the 2014 AISTATS conference). Our approach is based on the following key result that we proved: if an MLN has no shared terms then MAP inference over it can be reduced to MAP inference over a Markov network having the following properties: (i) the number of random variables in the Markov network is equal to the number of first-order atoms in the MLN; and (ii) the domain size of each variable in the Markov network is equal to the number of groundings of the corresponding first-order atom. This represents exponential complexity reductions over ground MAP inference. We improved this result further by showing that if non-shared MLNs contain no self joins, namely every atom appears at most once in each of its formulas, then all variables in the corresponding Markov network need only be bi-valued. .

Our approach is quite general and can be easily applied to arbitrary MLNs by simply grounding all of its shared terms. The key feature of our approach is that because we reduce lifted inference to propositional inference, we can use any propositional MAP inference algorithm for performing lifted MAP inference, without actually lifting the propositional algorithm. Within our approach, we experimented with two propositional MAP inference algorithms: Gurobi (an Integer Linear Programming solver) and MaxWalkSAT. Our experiments on several benchmark MLNs clearly demonstrated the superiority of our approach over the ground inference in terms of both the scalability and the solution quality.

PROBABILISTIC INFERENCE FOR HYBRID MLNS

General abductive inference requires the ability to reason about both discrete and continuous information. Accordingly, we have continued to develop an efficient algorithm for continuous, nonconvex optimization, which we call RDIS. RDIS allows us to efficiently perform MAP inference in continuous domains and to accurately fit models with continuous parameters to data. RDIS easily supports discrete variables as well as continuous, enabling it to play an important role in multi-modal and multi-scale domains. The key to RDIS is to identify and exploit local structure in the objective function by dynamically identifying a subset of variables that, once optimized, decomposes the remaining variables into approximately independent subsets. These can then be separately optimized, and we do so recursively in order to exploit multiple levels of local structure, dynamically finding within each a subset that further decomposes it, etc. RDIS is similar in structure to existing combinatorial algorithms and we accordingly use ideas from these and dynamically choose variables using hypergraph partitioning. This ensures that decomposition is achieved at each level, if at all possible. For value selection, RDIS employs ideas from continuous optimization in the form of a local optimization subroutine such as gradient descent or quasi-Newton to ensure that it can find the global optimum without needing to explore the infinitely-many values in the continuous domain. We've evaluated RDIS both analytically and empirically. Analytically, we show that RDIS finds the global minimum in exponentially less time than standard methods for a class of nonconvex functions that exhibit local structure. Empirically, tests on highly multi-modal functions, structure from motion, and protein folding demonstrate that our algorithm consistently finds significantly better minima than these same standard methods in challenging optimization and inference tasks.

TRACTABLE PROBABILISTIC KNOWLEDGE BASES (TPKB)

Tractable Markov Logic was originally designed to be used with Probabilistic Theorem Proving (PTP) as an inference algorithm, but PTP was much more complicated than necessary for the restricted case, which clouded intuition and made formal guarantees difficult. Last year, we worked on improvements of TML that feature existence uncertainty and an object oriented syntax. This year, we further improved TPKBs and learned a large-scale TPKB from a variety of data sources, a longstanding goal of AI research. While our large-scale probabilistic knowledge base is not the first large scale representation of such data, existing approaches either ignore the uncertainty inherent to knowledge extracted from text, the web, and other sources, or lack a consistent probabilistic semantics with tractable inference. TPKBs consist of a hierarchy of classes of objects and a hierarchy of classes of object pairs such that attributes and relations are independent conditioned on those classes. These characteristics facilitate both tractable probabilistic reasoning and tractable maximum-likelihood parameter learning. TPKBs feature a rich query language that allows one to express and infer complex relationships between classes, relations, objects, and their attributes. For instance, our query language features unions of existentially quantified conjunctive queries. These queries are translated to sequences of operations in a relational database facilitating query execution times in the sub-second range. We demonstrate the power of TPKBs by leveraging large data sets extracted from Wikipedia to learn their structure and parameters. The resulting TPKB models a distribution over millions of objects and billions of parameters. We apply the TPKB to entity resolution and object linking problems and show that the TPKB can accurately align large knowledge bases and integrate triples from open IE projects.

ALCHEMY 2.0 & ALCHEMY LITE

We continued to work on Alchemy (our open source software package for learning and inference in Markov Logic). The main highlight of Alchemy 2.0 is access to more scalable lifted inference algorithms such as lifted sampling approaches. We have also continued to work on algorithms that allow us to apply lifted inference algorithms to models that are usually not liftable. We also continued to develop Alchemy Lite, an open-source software package for inference in Tractable Markov Logic (TML), the first tractable first-order probabilistic logic. TML strikes a good balance between expressiveness and tractability, subsuming essentially all tractable models, including many high-treewidth ones. The software allows users to build intuitive models in an object-oriented style while guaranteeing that inference will be efficient without resorting to approximation or ad hoc performance hacks. Alchemy Lite allows for fast, exact inference for models formulated in terms of TML, as well as the ability to update models with new information. Further improvements to the inference implementation to allow for tasks such as entity resolution and parsing have also been under development.
For an extensive list of industry applications we refer the reader to the end of the report.

PROGRESS ON LEARNING

IMPROVED STRUCTURE LEARNING FOR SUM-PRODUCT NETWORKS

Learning the structure of SPNs is important for applying them to domains where the structure of the relationships among the variables is not known in advance. Our previous state-of-the-art algorithm (Gens & Domingos, 2013) for learning SPN structure performed top-down clustering of training instances and variables to create sum and product nodes, respectively. Variable

interactions are represented indirectly through sum nodes, which act as latent variables. In contrast, most algorithms for learning graphical models represent interactions directly through conditional probability distributions or potential functions, not through latent variables.
We developed a new SPN structure learning algorithm, called ID-SPN, for learning SPNs with both indirect and direct variable interactions. ID-SPN performs a top-down clustering of instances and variables, similar to our previous method, but with tractable graphical models at the leaves instead of univariate distributions. These leaf distributions are learned using ACMN, a state-of-the-art method we developed for learning arithmetic circuits (Lowd & Rooshenas, 2013). ID-SPN uses the likelihood of a validation set to determine the proper depth of each branch. In experiments on a standard set of benchmark datasets, ID-SPN is more accurate than the previous state-of-the-art on 20 out of 20 datasets. ID-SPN also obtains better likelihoods than intractable Bayesian networks on 13 out of 20 datasets, suggesting that tractable models can be as accurate as intractable ones.

DEEP SYMMETRY NETWORKS

The chief difficulty in object recognition is that objects' classes are obscured by a large number of extraneous sources of variability, such as pose and part deformation. These sources of variation can be represented by symmetry groups, sets of composable transformations that preserve object identity. Our goal is to combine the rich tractable inference of Sum-Product Networks with the generalization of symmetry groups. Deep Symmetry Networks show the benefit of extending neural networks to richer symmetry spaces. Convolutional neural networks (convnets) achieve a degree of translational invariance by computing feature maps over the translation group, but cannot handle other groups. As a result, these groups' effects have to be approximated by small translations, which often requires augmenting datasets and leads to high sample complexity. We developed deep symmetry networks (symnets), a generalization of convnets that forms feature maps over arbitrary symmetry groups. Symnets use kernel-based interpolation to tractably tie parameters and pool over symmetry spaces of any dimension. They also use a Lucas-Kanade optimization to warp features to feature maps. These techniques sidestep the exponential computational burden of convolving a feature in high dimensions. Like convnets, symnets are trained with backpropagation. The composition of feature transformations through the layers of a symnet provides a new approach to deep learning. Our preliminary experiments on the NORB and MNIST-rot datasets show that symnets over the affine group greatly reduce sample complexity relative to convnets by better capturing the symmetries in the data.

SYMMETRY-BASED SEMANTIC PARSING

An abductive inference system should be capable of interpreting all kinds of information. Many times, information will come into the system in the form of unstructured text. Typically, the strategy for dealing with unstructured text is use a semantic parser to map the text to its formal meaning representation in some first-order logic language. However, there is little consensus about the best meaning representation to choose and finding enough labeled training data to train such a semantic parser is difficult and costly. We have proposed a new notion of semantics that avoids these challenges, and, as added benefit, represents meaning in a way that is more easily integrated into a Tractable Probabilistic Knowledge Base (TPKB). We utilize insights from symmetry group theory, which studies the formal properties of symmetry groups, which are groups of transformations under which key properties of a structure are preserved. We introduce the concept of a semantic symmetry group, which contains syntactic operations which when applied to a sentence preserve its meaning. A semantic symmetry group partitions the set of all sentences into sets, called orbits, of sentences with the same meaning. The orbit that a sentence is a member of implicitly defines its meaning. Since natural language frequently contains ambiguities, we utilize a probabilistic approach to semantic symmetry and orbit membership. Properties of symmetry group theory allow the design of compact probabilistic models of meaning over which inference is efficient and that can be assimilated into a TPKB. We have begun implementing a symmetry-based semantic parser that will learn a semantic symmetry group in an unsupervised way from text.

EXCHANGEABLE VARIABLE MODELS

A sequence of random variables is exchangeable if its joint distribution is invariant under variable permutations. We introduce exchangeable variable models (EVMs) as a novel class of probabilistic models whose basic building blocks are partially exchangeable sequences, a generalization of exchangeable sequences. We present conditions that imply tractable probabilistic inference and discuss parameter and structure learning. We prove that a family of tractable EVMs is optimal under zero-one loss for a large class of functions, including parity and threshold functions, and strictly subsumes existing tractable independence-based model families. Extensive experiments show that EVMs outperform state of the art classifiers such as SVMs and probabilistic models which are solely based on independence assumptions.

This year, we proposed exchangeable variable models (EVMs), a novel family of probabilistic models for classification and probability estimation. While most probabilistic models are built on the notion of conditional independence and its graphical representation, EVMs have finite partially exchangeable sequences as basic components. We show that EVMs can represent complex positive and negative correlations between large sets of variables with few parameters and without sacrificing tractable inference. The parameters of EVMs are estimated under the maximum-likelihood principle and we assume the examples to be independent and identically distributed. We develop methods for efficient probabilistic inference, maximum-likelihood estimation, and structure learning.

We also introduced the mixtures of EVMs (MEVMs) family of models which is strictly more expressive than the naive Bayes family of models but as efficient to learn. MEVMs represent classifiers that are optimal under zero-one loss for a large class of Boolean functions including parity and threshold functions. Extensive experiments show that exchangeable variable models, when combined with the notion of conditional independence, are effective both for classification and probability estimation. The MEVM classifier significantly outperforms state of the art classifiers on numerous high-dimensional and sparse data sets. MEVMs also outperform several tractable graphical model classes on typical probability estimation problems while being orders of magnitudes more efficient.

RELATIONAL SUM-PRODUCT NETWORKS

Relational Sum-Product Networks: Sum-product networks (SPNs; Poon and Domingos 2011) are a recently-proposed deep architecture that guarantees tractable inference, even on certain high-treewidth models. SPNs are a propositional architecture, treating the instances as independent and identically distributed. Last year, we developed Relational Sum-Product Networks (RSPNs), a new tractable first-order probabilistic architecture. RSPNs generalize SPNs by modeling a set of instances jointly, allowing them to influence each other's probability distributions, as well as modeling the probabilities of relations between objects. This year, we developed LearnRSPN, the first algorithm for learning high-treewidth tractable statistical relational models. LearnRSPN is a recursive top-down structure learning algorithm for RSPNs, based on the LearnSPN (Gens and Domingos 2013) algorithm for propositional SPN learning. In our empirical evaluation, the RSPN learning algorithm outperforms Markov Logic Networks (Richardson and Domingos 2006) in both running time and predictive accuracy.

TRACTABLE PROBABILISTIC PROGRAMS

We also developed Tractable Probabilistic Programs (TPP), a statistical relational representation that captures a probability distribution over programs drawn from a context-free grammar. Although TPP was motivated by the problem of automated software debugging, we are investigating its applicability to other problem domains that make use of probabilistic grammars, such as natural language processing and image understanding.

LEARNING CUTSET NETWORKS

Learning tractable probabilistic models from data has been the subject of much recent research. These models offer a clear advantage over Bayesian networks and Markov networks: exact inference over them can be performed in polynomial time, obviating the need for unreliable, inaccurate approximate inference, not only at learning time but also at query time. Interestingly, experimental results in numerous recent studies have shown that the performance of approaches that learn tractable models from data is similar or better than approaches that learn Bayesian and Markov networks from data. These results suggest that controlling exact inference complexity is the key to superior end-to-end performance.

To take advantage of these promising results, in a recent work that will appear in ECML 2014, we introduced a new tractable probabilistic model called cutset networks for representing large, multi-dimensional discrete distributions. Cutset networks are rooted OR search trees, in which each OR node represents conditioning of a variable in the model, with tree Bayesian networks (Chow-Liu trees) at the leaves. From an inference point of view, cutset networks model the mechanics of Pearl's cutset conditioning algorithm, a popular exact inference method for probabilistic graphical models. We developed efficient algorithms, which leverage and adopt vast amount of research on decision tree induction for learning cutset networks from data. We also developed an expectation-maximization (EM) algorithm for learning mixtures of cutset networks. Our experiments on a wide variety of benchmark datasets clearly demonstrated that compared to approaches for learning other tractable models such as thin-junction trees, latent tree models, arithmetic circuits and sum-product networks, our approach is significantly more scalable, and provides similar or better accuracy on all datasets. In fact, it was better than the competing seven state-of-the-art algorithm on 55\% of the datasets. We are quite excited about these results because the algorithm is quite fast (has provably small computational complexity) and achieves state-of-the-art performance. This makes it an ideal candidate for learning in “Big data” domains.

COMBINING MARKOV LOGIC AND SUPPORT VECTOR MACHINES FOR EVENT EXTRACTION

Event extraction is the task of extracting and labeling all instances in a text document that correspond to a predefined event type. This task is quite challenging because of a multitude of reasons: events are often nested, recursive and have several arguments; there is no clear distinction between arguments and events; etc. For instance, consider the BioNLP Genia shared task Nedellec et al. (2013). In this task, participants are asked to extract instances of a predefined set of Biomedical events from text. An event can have an arbitrary number of arguments that correspond to predefined argument types, and is identified by a keyword called the trigger. The task is complicated by the fact that an event may serve as an argument of another event (nested events). We made the following contributions.

First, we proposed a novel model for Biomedical event extraction based on MLNs that leverages the power of support vector machines (SVMs) Joachims (1999); Vapnik (1995) to handle high-dimensional features. Specifically, we (1) learned SVM models using rich linguistic features for trigger and argument detection and type labeling; (2) designed an MLN composed of soft formulas (each of which encodes a soft constraint whose associated weight indicates how important it is to satisfy the constraint) and hard formulas (constraints that always need to be satisfied, thus having a weight of 1) to capture the relational dependencies between triggers and arguments; and (3) encoded the SVM output as prior knowledge in the MLN in the form of soft formulas, whose weights are computed using the confidence values generated by the SVMs. This formulation naturally allows SVMs and MLNs to complement each other's strengths and weaknesses: learning in a large and sparse feature space is much easier with SVMs than with MLNs, whereas modeling relational dependencies is much easier with MLNs than with SVMs.

Our second contribution concerns making inference with this MLN feasible. Recall that inference involves detecting and assigning the type label to all the triggers and arguments. We showed that existing Maximum-a-posteriori (MAP) inference methods, even the most advanced approximate ones (e.g., Selman et al. (1996), Sontag and Globerson (2011), Marinescu and Dechter (2009), etc.), are infeasible on our proposed MLN because of their high memory cost. To combat this, we identified decompositions of the MLN into disconnected components and solved each independently, thereby drastically reducing the memory requirements.

We evaluated our approach on the BioNLP 2009, 2011 and 2013 Genia shared task datasets. On the BioNLP'13 dataset, our model significantly outperforms state-of-the-art pipeline approaches and achieves the best F1 score to date. On the BioNLP'11 and BioNLP'09 datasets, our scores are slightly better and slightly worse respectively than the best reported results. However, they are significantly better than state-of-the-art MLN-based systems. A paper on this work will appear at the 2014 Empirical Methods in Natural Language Processing (EMNLP) conference.

LEARNING BAYESIAN LOGIC PROGRAMS FOR TEXTUAL INFERENCE

The aim of this on-going part of the project is to automatically learn Bayesian Logic Programs (BLPs) from information extracted from natural-language text and use the resulting probabilistic model to make accurate "abductive" inferences from facts extracted from future documents.

We participated in the NIST KBP (Knowledge-Based Population) slot-filling task by using a BLP developed for the KBP ontology to make inferences from text extractions with the goal of increasing recall. We used the publicly distributed version of the CUNY BLENDER system as the base-level KBP extractor. During testing, we used a learned BLP to infer additional facts from the facts extracted by BLENDER, and submitted two sets of results for the competition, one with inferred relations added as well as a baseline set of results without BLP inferences. In order to assemble a large training set for learning a BLP appropriate for KBP, we mapped 26 of the 41 predicates in the KBP ontology to relations in the open-linked database, DBPedia. We then used our previously developed on-line BLP rule learner to learn a BLP from 912,375 mapped facts from DBpedia. For example, one learned rule was: "If person B is a key employee of organization A, then B is probably a shareholder in A." Unfortunately, partly because the KBP evaluation is focused on evaluating the extraction of explicitly-stated facts rather than probable inferences, the BLP inferences failed to improve recall and actually resulted in an overall decrease in F-measure (form 0.123 to 0.108). In our officially submitted results, we preferred inferred slot fillers to explicitly extracted ones in order to emphasize the role of inference. Subsequent to the official evaluation, we conducted an additional experiment in which we preferred inferred fillers to extracted ones only if their estimated confidence was higher. This version generated 7 additional fillers that were judged correct, resulting in an increase in recall (from .079 to .085) with only a minor decrease in F-measure (from 0.123 to 0.121). This result provides evidence for the value of BLP textual inference despite the limitations of the KBP evaluation with respect to evaluating this capability.

Our recent work has focused on scaling our BLP learning and inference methods to large-scale linked open data, specifically DBPedia. The goal is to learn a BLP from such large-scale data, map the ontology to that for a particular text-extraction task (e.g. KBP), and then use the BLP to make inferences from initial information extracted from text. In order to scale BLP learning to large multi-relational databases such as DBPedia, we have adapted the rule-learning algorithm of Ni Lao at al. (EMNLP, 2011, 2012) to learn the initial relational rules. We then use a simple, approximate maximum-likelihood parameter-learning method we have developed for conditional probability tables (CPTs) that use noisy-or and noisy-and to learn a BLP based on these rules. In order to scale inference to the large, complex BLPs learned from such data, we exploit a semantic-web-based implementation of DataLog, called JENA, to support logical inference, and Gogate's SampleSearch method for efficient and effective probabilistic inference for graphical models with both deterministic and probabilistic constraints.

We have also finalized our plans for evaluating the learned BLPs using the data available in DBPedia, and are currently in the process of conducting a full-scale experimental evaluation. We are using cross-validation on DBPedia data to directly evaluate to the accuracy of BLP-derived inferences. For each fact in a subset of DBPedia, we delete the fact from the database and attempt to infer a value for the corresponding slot using the learned BLP. For example, if we delete the fact that Natasha Obama is a child of Barack Obama, we may be able to infer it from the fact that Natasha is Malia Obama's sister and that Malia is a child of Barack Obama. By using the probability computed using the BLP model to rank the inferred fillers of a slot, we are generating an average precision-recall curve and computing the Mean Average Precision (MAP) to evaluate the accuracy of inference. By comparing the results of BLP inference to that of a purely logical approach (which is unable to meaningfully rank inferred fillers), we plan to measure the advantage of the BLP approach. Preliminary experiments using this methodology have demonstrated promising results, and we are in the process of completing comprehensive experiments using cross-validation on DBPedia. This work will continue as part of the DARPA DEFT project.

DISTRIBUTIONAL MARKOV LOGIC SEMANTICS

The goal of this aspect of the project is to develop an approach to representing the meaning of natural-language sentences as rich, formal expressions in probabilistic logic. An initial logical form is obtained by parsing a sentence using Combinatory Categorial Grammar (CCG). Next, uncertain, distributional information is added as weighted inference rules. The result is a "deep" representation of semantics that captures both logical structure as well as probabilistic, distributional meaning of words and phrases. This representation then supports rich "abductive" probabilistic inference from natural-language text using both Markov logic and Probabilistic Soft Logic (PSL). In particular, we have evaluated the approach on two standard textual inference problems, Recognizing Textual Entailment (RTE) and Semantic Textual Similarity (STS).

Recently, we have worked on improving the efficiency and accuracy of MLN inference for natural-language semantics. We have also explored the use of Probabilistic Soft Logic (PSL) for the STS task.

In March 2014, we participated in Task 1 of SemEval (Semantic Evaluation Workshop): "Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and entailment". The task involved both RTE and STS subtasks on the SICK dataset

(Sentences Involving Compositional Knowledge, Marelli et al., to appear). We obtained an accuracy of 73% on RTE, and a Pearson correlation of 0.71 on STS.

Markov Logic Networks can handle all of first-order logic, and have a principled basis in probabilistic logic; however, the networks can grow very large, leading to intractable inference. We have integrated a new inference algorithm based on SampleSearch into Alchemy (the MLN inference system that we are using) to improve run time. We also introduced a modified closed-world assumption that significantly reduces the size of the ground network, thereby making inference feasible. This step has the added benefit of removing extraneous literals from the system, thereby making inference more accurate. Evaluation on the training portion of the SICK RTE data yielded an accuracy of 71.8% for the modified system (original system: 56.9%) with an average runtime of 7s per datapoint (original system: 2min 27s).

We have also explored Probabilistic Soft Logic (Boecheler, Mihalkova and Getoor 2010) as an alternative framework for probabilistic inference for the STS task. We changed the interpretation function for conjunction in PSL to a weighted average to make it more appropriate for STS. In addition, we implemented a new heuristic variant of the lazy grounding implemented in PSL designed to work with the changed implementation of conjunction in a way that avoids the construction of irrelevant groundings. We obtained Pearson correlations of 0.79 on the MSR video corpus, 0.53 on the MRS paraphrase corpus, and 0.71 on the training portion of the SICK STS dataset. In addition, inference was an order of magnitude faster with PSL than with MLNs.

The SICK RTE data allows for three judgments on whether the Text (T) entails the Query (Q): either Entailment, Contradiction, or Neutral. In order to model this three-way distinction, we computed two probabilities, P(Q|T) and P(Q|not(T)), and used a supervised classifier to choose a judgment based on these two probabilities. This setup has the added benefit of addressing the fundamental problem of MLNs that the computed probability of a sentence depends on both the domain size and the size of the sentence.

Our model combines deep semantics through logical form with weighted inference rules derived from distributional models and can be viewed as an approach to Semantic Parsing that, instead of using a fixed, manually created ontology to interpret predicates, interprets predicate symbols using distributional rules that are automatically created "on the fly."

ROBUST STRUCTURED PREDICTION THROUGH REGULARIZATION

In previous work, we developed max-margin learning methods for collective classification that are robust to adversarial manipulation of object features (Torkamani & Lowd, 2013). However, these methods were restricted to associative Markov networks and could not handle more complex scenarios, such as adversaries that manipulate link structure. We developed a new strategy for learning robust Markov networks or structural SVMs by showing that robustness to perturbations of the features is equivalent to regularization. Specifically, when perturbations are constrained by a norm, the equivalent regularizer is given by the dual norm. When perturbations are constrained by a polyhedron, the equivalent regularizer is a linear function in a transformed space. In experiments, we demonstrate that this regularization strategy leads to improved generalization on a collective classification problem with a lot of concept drift.

DISTRIBUTED GRAPHLAB

While high-level data parallel frameworks, like MapReduce, simplify the design and implementation of large-scale data processing systems, they do not naturally or efficiently support many important data mining and machine learning algorithms and can lead to inefficient learning systems. To help fill this critical void, we introduced the GraphLab abstraction which naturally expresses asynchronous, dynamic, graph-parallel computation while ensuring data consistency and achieving a high degree of parallel performance in the shared-memory setting. We extended the GraphLab framework to the substantially more challenging distributed setting while preserving strong data consistency guarantees. We developed graph based extensions to pipelined locking and data versioning to reduce network congestion and mitigate the effect of network latency. We also introduced fault tolerance to the GraphLab abstraction using the classical Chandy-Lamport snapshot algorithm and demonstrate how it can be easily implemented by exploiting the GraphLab abstraction itself. Finally, we evaluated our distributed implementation of the GraphLab abstraction on a large Amazon EC2 deployment and show 1-2 orders of magnitude performance gains over Hadoop-based implementations.

PARALLEL LEARNING FOR GRAPHICAL MODELS

Two strong areas of focus have been in graphical models and parallel learning. To address these problems in a more accurate fashion, we've developed a gradient boosting algorithm for tree-shaped conditional random fields (CRF). Conditional random fields are an important class of models for accurate structured prediction, but effective design of the feature functions is a major challenge when applying CRF models to real world data. Gradient boosting, which can induce and select functions, is a natural candidate solution for the problem. However, it is non-trivial to derive gradient boosting algorithms for CRFs, due to the dense Hessian matrices introduced by variable dependencies. We address this challenge by deriving a Markov Chain mixing rate bound to quantify the dependencies, and introduce a gradient boosting algorithm that iteratively optimizes an adaptive upper bound of the objective function. The resulting algorithm induces and selects features for CRFs via functional space optimization, with provable convergence guarantees. Experimental results on three real world datasets demonstrate that the mixing rate based upper bound is effective for training CRFs with non-linear potentials.

GRAPHLAB: CODE RELEASE AND TECHNOLOGY TRANSFER

One of the major goals of this project is the development of open-source software and of a community around it. We have held two GraphLab workshops in the last couple of years. The first one in 2012 had 318 people in attendance. The second one in 2013 had 570 people. All our code is available at http://graphlab.org. The GraphLab open-source project, started by PI

Guestrin, has received very significant attention in industry and academia. As discussed above, the software has received tens of thousands of downloads, and held two very popular workshops. This project has had very significant impact in industry and academia. To continue to support the users of GraphLab, and to continue to expand its reach, we have recently spun off a company, where GraphLab can become its own entity beyond the university. This company has recently announced its first round of funding, receiving $6.75M.

PROGRESS ON ACTIVITY AND PLAN RECOGNITION

Activity recognition is central to many important problems including surveillance (recognizing the activities of an opponent), anomaly detection (recognizing and ignoring normal behaviors), and human-computer interaction (recognizing the goals and activities of the user). Activity recognition algorithms seek to infer the goals and plans of one or more agents from noisy and fragmentary observations of their behavior. This is a classic problem of abductive inference.

ACTIVITY RECOGNITION IN THE KITCHEN

In our first application, we implemented and evaluated a Markov-logic based plan recognition system for kitchen activities, where observations came from video and natural language narration (Song et al 2013). We presented a general framework for complex event recognition that is well-suited for integrating information that varies widely in detail and granularity. Consider the scenario of an agent in an instrumented space performing a complex task while describing what he is doing in a natural manner. The system takes in a variety of information, including objects and gestures recognized by RGB-D and descriptions of events extracted from recognized and parsed speech. The system outputs a complete reconstruction of the agent's plan, explaining actions in terms of more complex activities and filling in unobserved but necessary events. We show how to use Markov Logic (a probabilistic extension of first-order logic) to create a model in which observations can be partial, noisy, and refer to future or temporally ambiguous events; complex events are composed from simpler events in a manner that exposes their structure for inference and learning; and uncertainty is handled in a sound probabilistic manner.

We evaluated our framework on a multi-modal corpus collected from people conducting tasks in an instrumented kitchen, including making tea, making cocoa and making oatmeal. Participants were asked to conduct the activity and at the same time verbally describe the action being conducted. The experiments demonstrated that (i) employing a complex event library improves visual event detection, and (ii) using both an event library and data from free-form spoken language can compensate for sparse visual input.

ACTIVITY RECOGNITION IN SOCIAL MEDIA

We extended our work on activity and state recognition from social media data (Sadelik et al 2013, Brennan et al 2013). Computational approaches to health monitoring and epidemiology continue to evolve rapidly. We presented an end-to-end system, nEmesis, that automatically identifies restaurants posing public health risks. Leveraging a language model of Twitter users" online communication, nEmesis finds individuals who are likely suffering from a foodborne illness. People"s visits to restaurants are modeled by matching GPS data embedded in the messages with restaurant addresses. As a result, we can assign each venue a "health score" based on the proportion of customers that fell ill shortly after visiting it. Statistical analysis reveals that our inferred health score correlates (r = 0:30) with the official inspection data from the Department of Health and Mental Hygiene (DOHMH). We investigated the joint associations of multiple factors mined from online data with the DOHMH violation scores and find that over 23% of variance can be explained by our factors. We demonstrated that readily accessible online data can be used to detect cases of foodborne illness in a timely manner. This approach offers an inexpensive way to enhance current methods to monitor food safety (e.g., adaptive inspections) and identify potentially problematic venues in near-real time.

MAKING MODEL MARKOV LOGIC MORE EFFICIENT

We developed complexity results on a "less intractable" subset of multi-agent Markov Logic (Papai & Kautz 2013). Modal Markov Logic for a single agent has previously been proposed as an extension to propositional Markov logic. While the framework allowed reasoning under the principle of maximum entropy for various modal logics, it is not feasible to apply its counting based inference to reason about the beliefs and knowledge of multiple agents due to magnitude of the numbers involved. We propose a modal extension of propositional Markov logic that avoids this problem by coarsening the state space. The problem stems from the fact that in the single-agent setting, the state space is only doubly exponential in the number of propositions in the domain, but the state space can potentially become infinite in the multi-agent setting. In addition, the proposed framework adds only the overhead of deciding satisfiability for the chosen modal logic on the top of the complexity of exact inference in propositional Markov logic. The proposed framework allows one to find a distribution that matches probabilities of formulas obtained from training data (or provided by an expert). Finally, we showed how one can compute lower and upper bounds on probabilities of arbitrary formulas.

PLAN RECOGNITION WITH MONTE CARLO TREE SEARCH

We continued investigating the application of Monte Carlo tree search (MCTS algorithms to planning and plan recognition. This year, we developed the mathematical foundations for state abstraction in MCTS. In previous work, we pioneered formal methods for temporal and state abstraction in hierarchical reinforcement learning (HRL). The requirements for correct state abstraction in HRL are very stringent and, hence, rarely satisfied in practice. In contrast, we showed that state abstraction in MCTS is much easier to achieve. We proved accuracy bounds for a certain form of state abstraction, state aggregation in ExpectiMax trees, and we showed that these state abstractions preserve optimality in search trees. This in turn permitted us to prove correctness of state aggregation abstractions for two MCTS methods: Sparse Sampling and UCT. These results are very general and show excellent performance improvements in several benchmark problems. Future work will focus on automatically learning these state abstractions.

TECHNOLOGY TRANSFER

The software and methods developed as part of the project have found numerous applications both in industry and academia. In the following, we list some of these technology transfers.

- DARPA PPAML (Probabilistic Programming for Advanced Machine Learning):
    - The project was launched in Fall 2013, and developed in part based on research funded under this MURI in the Tenenbaum group. An additional MURI member (Dietterich) is playing a crucial role on the PPAML evaluation team.
    - Several teams of the project are using algorithms in Alchemy 2.0 for building their probabilistic programming systems as well as for competing in the DARPA evaluations.
- Vibhav Gogate along with Avi Pfeffer from Charles Rivers Analytics received a Phase 1 AFOSR SBIR grant on “Representation and Inference for Developing Deep Language Engines (RIDDLE).” The project used lifted algorithm in Alchemy 2.0 for solving complex NLP tasks such as Event Extraction and temporal relation classification.
- Daniel Lowd received a Google faculty research award funding work on tractable probabilistic models, research that was initiated as part of the MURI project.
- Tractable Markov logic and Tractable Probabilistic Knowledge Bases are applied and extended by Domingos’ group within the DARPA DEFT (Deep Exploration of Text) project and the ONR BRC project on Structured Learning for Scene Understanding.
- Tenenbaum helped to organize and keynote a workshop on the interface between computational cognitive modeling and the Intelligence Community, sponsored by IARPA and BBN/Raytheon. This workshop was attended by approximately 80 members of the IC and associated research agencies, at Raytheon offices near Fort Meade. MURI funded research was presented and generated significant interest and follow-up from multiple attendees.
- Core research from this grant (the GraphLab system) was spun off as a company. This start up received $6.75M in VC funding, and is currently employing 26 people.

COMPANIES AND INDIVIDUALS WHO HAVE WORKED WITH ALCHEMY

- Google Inc
    - Kevin Murphy (murphyk@cs.ubc.ca)
    - Brian Milch (milch@google.com)
- Microsoft Research
    - Past Contributors to Alchemy already at Microsoft research (Matt Richardson, Hoifung Poon)
    - Ben Livshits (livshits@microsoft.com);
    - many others
- IBM Research
    - Ashish Sabharwal (now at AI2; Paul Allen Institute in Seattle)
- LogicBlox
    - Benny Kimelfeld (bennyk@gmail.com)
    - Molham Aref (molham.aref@logicblox.com)
- Charles Rivers Analytics
    - Avi Pfeffer (apfeffer@cra.com)
- Facebook
    - Kedar Bellare (kedar.bellare@gmail.com)
- Baylor Hospital Network
    - Dr. Jack Stecher (jackstecher@sbcglobal.net)
- Amazon
    - Srikanth Doss (srikanth.doss@gmail.com)
- Netflix
    - Qi Zhong (no longer at Netflix)
- Yahoo
    - Andrew Gelfand (agelfand@ics.uci.edu)
    - Vibhor Rastogi (no longer works at Yahoo; now at Facebook)
- BAE systems
    - Gregory Sullivan (gregory.sullivan@baesystems.com)
- Raytheon
    - Kenric P Nelson (Kenric_P_Nelson@raytheon.com)
- SRI International
    - Rodrigo de Salvo Braz (braz@ai.sri.com)
- Samsung
    - Jae Hyun Son (jhson@sta.samsung.com)
- Match.com
    - Vaclav Patricek (patricek@match.com)
- BBN technology
    - Hala Mostafa (hmostafa@bbn.com)
- CycCorp
    - Michael Witbrock (witbrock@cyc.com)
- Gamelan systems
    - Ben Vigoda (ben.vigoda@gamelanlabs.com)
- Analog Devices

    - Theo Weber (theo.weber@analog.com)
- Intel
    - Denver Dash (now a research Professor at cmu)
    - Many others affiliated with center at UW
- Nuance Communications
    - Hung Bui (bui.h.hung@gmail.com)
- Ebay
    - Ravi Jammalamadaka (ravij@ebay.com)
- Hitachi
    - Robert Mateescu (mateescu@hitachi.com)
- Other companies which have used Alchemy but we do not have contacts for
    - At&T, Nokia, Twitter, Xerox Corp

REFERENCES

Bahar, R., Frohm, E., Gaona, C., Hachtel, G., Macii, E., Pardo, A., and Somenzi, F. (1993). Algebraic decision diagrams and their applications. In Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, pages 188-191.

Bonawitz, E. and Ullman, T. and Gopnik, A. and Tenenbaum, J.B. (2012). Sticking to the evidence? A computational and behavioral case study of micro-theory change in the domain of magnetism. Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics.

Cheng, J. and Druzdzel, M. J. (2000). AIS-BN: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. Journal of Artificial Intelligence Research, 13:155-188.

Chol Song, Y. and Kautz, H. (2012). A Testbed for Learning by Demonstration from Natural Language and RGB-Depth Video. Proceedings of the 26th Conference on Artificial Intelligence.

Darwiche, A. (2003). A Differential Approach to Inference in Bayesian Networks. Journal of the ACM, 50:280-305.

Diaconis, Persi and Freedman, David. De Finetti's generalizations of exchangeability. In Studies in Inductive Logic and Probability, volume II. 1980a.

Diaconis, Persi and Freedman, David. Finite exchangeable sequences. The Annals of Probability, 8(4):745–764, 1980b.

Dechter, E. and Malmaud, J. and Adams, R.P. and Tenenbaum, J.B. (2013). Bootstrap learning via modular concept discover. Proceedings of the International Joint Conference on Artificial Intelligence.

Domingos, P. and Lowd, D. (2009). Markov Logic: An Interface Layer for Artificial Intelligence. Morgan & Claypool, San Rafael, CA.

Domingos, P. and Webb, A. (2012) A Tractable First-Order Probabilistic Logic. Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.

Domingos, P. and Webb, A. (2012). A tractable First-order probabilistic logic. Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.

Duvenaud, D. and Lloyd, J.R. and Grosse, R. and Tenenbaum, J.B. and Ghahramani, Z. (2013). Structure Discovery in Nonparametric Regression through Compositional Kernel Search. Proceedings of the International Conference on Machine Learning.

Elidan, G. and Globerson, A. (2010). The 2010 UAI approximate inference challenge. Available online at: http://www.cs.huji.ac.il/project/UAI10/index.php.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721-741.

Gens, R. and Domingos, P. (2012) Discriminative Learning of Sum-Product Networks.

Advances in Neural Information Processing Systems 25.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). Markov Chain Monte Carlo in Practice. Chapman and Hall, London, UK.

Gogate, V. and Dechter, R. (2007). SampleSearch: A scheme that Searches for Consistent Samples. Proceedings of the 11th Conference on Artificial Intelligence and Statistics (AISTATS), pages 147-154.

Gogate, V. and Dechter, R. (2011). SampleSearch: Importance sampling in presence of determinism. Artificial Intelligence, 175(2):694-729.

Gogate, V. and Domingos, P. (2011). Probabilistic Theorem Proving. Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence.

Gonzalez, J. and Low, Y. and Gu, H. and Bickson, D. and Guestrin, C. (2012). PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs. Proceedings of Operating Systems Design and Implementation.

Grosse, R.B. and Salakhutdinov, R. and Freeman, W.T. and Tenenbaum, J.B. (2012). Exploiting compositionality to explore a large space of model structures. Proceedings of the Conference on Uncertainty in Artificial Intelligence.

Jara-Ettinger, J. and Baker, C.L. and Tenenbaum, J.B. (2012). Learning What is Where from Social Observations. Proceedings of the 34th Annual Conference of the Cognitive Science Society.

Jensen, C. S., Kjaerulff, U., and Kong, A. (1993). Blocking Gibbs Sampling in Very Large Probabilistic Expert Systems. International Journal of Human Computer Studies. Special Issue on Real-World Applications of Uncertain Reasoning, 42:647-666.

Kyrola, A., and Blelloch, G. and Guestrin, C. (2012). GraphChi: Large-Scale Graph computation on Just a PC. Proceedings of the 10th USENIX Symposium on Operating Systems Design and Implementation.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. Journal of the Royal Statistical Society. Series B (Methodological), 50(2):157-224.

Liu, J. S., Wong, W. H., and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes. Biometrika, 81:27-40.

Low, Y. and Gonzalez, J. and Kyrola, A. and Bickson, D. and Guestrin, C. and Hellerstein, J.M.

(2012). Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud. Proceedings of Very Large Data Bases.

Swift, M. and Ferguson, G. and Galescu, L., and Chu, Y. and Harman, C. and Jung, H. and Perera I. and Chol Song, Y. and Allen J. and Kautz, H. (2012). A multimodal corpus for integrated language and action. Proceedings of the International Workshop on Multimodal Corpora for Machine Learning.

Poon, H and Domingos, P. (2011) Sum-Product Networks: A New Deep Architecture. Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence.

Sarkhel, S. and Gogate, V. (2013). Lifting walksat-based local search algorithms for map inference. Proceedings of the 3rd International Workshop on Statistical Relational Artificial Intelligence.

Smith, D. and Gogate, V. (2013). The inclusion-exclusion rule and its application to the junction tree algorithm. Proceedings of the International Joint Conference on Artificial Intelligence.

Venugopal, D. and Gogate, V. (2013a). Dynamic blocking and collapsing for gibbs sampling. Proceedings of the Conference on Uncertainty in Artificial Intelligence.

Venugopal, D. and Gogate, V. (2013b). Giss: Combining SampleSearch and importance sampling for inference in mixed probabilistic and deterministic graphical models. Proceedings of the AAAI Conference on Artificial Intelligence.

Webb, W.A. and Domingos, P. (2013). Tractable Probabilistic Knowledge Bases with Existence Uncertainty. Proceedings of the 3rd International Workshop on Statistical Relational Artificial Intelligence.